Mapping the receptivity of malaria risk to plan the future of control in Somalia

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2012-001160
Article Type:	Research
Date Submitted by the Author:	13-Mar-2012
Complete List of Authors:	Noor, Abdisalan; Kenyan Medical Researh Institute/University of Oxford-Wellcome Trust Research Programme, Malaria Public Health & Epidemiology Group Alegana, Victor; Kenyan Medical Researh Institute/Wellcome Trust Research Prgramme, Malaria Public Health & Epidemiology Group, Nairobi, Kenya Patil, Anand; Sense Inc., Moloney, Grainne; United Nations Food and Agricultural Organization, Food Security and Nutrition Analysis Unit-Somalia Borle, Mohammed; United Nations Food and Agricultural Organization, Food Security and Nutrition Analysis Unit-Somalia Yusuf, Fahmi; World Health Organization, Roll Back Malaria, Somalia Office Amran, Jamal; World Health Organization, Roll Back Malaria, Somalia Office Snow, Robert; Kenyan Medical Researh Institute/University of Oxford-Wellcome Trust Research Programme, Malaria Public Health & Epidemiology Group
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Public health, Infectious diseases, Health policy
Keywords:	Epidemiology < INFECTIOUS DISEASES, Public health < INFECTIOUS DISEASES, Geographical mapping < TROPICAL MEDICINE

Mapping the receptivity of malaria risk to plan the future of control in Somalia

Abdisalan M Noor [1,2*], Victor A Alegana[1], Anand P Patil[3], Grainne Moloney[4], Mohammed Borle[4], Fahmi Yusuf[5], Jamal Amran[5], Robert W Snow[1,2]

1. Malaria Public Health & Epidemiology Group, Kenya Medical Research Institute/Wellcome Trust Research Programme, Centre for Geographic Medicine Research-Coast, P.O. Box 43640, 00100 GPO, Nairobi, Kenya.
2. Nuffield Department of Medicine, Centre for Tropical Medicine, University of Oxford, John Radcliffe Hospital, Headington, Oxford, OX3 9DU, UK
3. Sense Inc., 90 Lawrence St., Detroit, MI, USA, 48202.
4. United Nations Food and Agricultural Organization, Food Security and Nutrition Analysis Unit-Somalia, 3rd Floor P.O. Box 1230, Village Market, Nairobi, Kenya
5. World Health Organization, Malaria Control & Elimination, Somalia.

***AMN:** anoor@nairobi.kemri-wellcome.org
VAA : valegana@nairobi.kemri-wellcome.org
APP : anand.prabhakar.patil@gmail.com
GM : Grainne.Moloney@fao.org
MB: Mohammed.Borle@fao.org
FY: yusuff@som.emro.who.int
JA : amranj@nbo.emro.who.int
RWS: rsnow@nairobi.kemri-wellcome.org

*Author for correspondence

ABSTRACT

Objectives: To measure the receptive risks of malaria in Somalia and compare decisions on intervention scale-up based on this map and the more widely used contemporary risk maps.

Design: Cross-sectional community *Plasmodium falciparum* parasite rate (*Pf*PR) data for the period 2007 to 2010 corrected to a standard age-range of two to less than 10 years ($PfPR_{2-10}$) and used within a Bayesian space-time geostatistical framework to predict the contemporary (2010) mean $PfPR_{2-10}$ and the maximum annual mean $PfPR_{2-10}$ (receptive) from the highest predicted $PfPR_{2-10}$ value over the study period as an estimate of receptivity.

Setting: Randomly sampled communities in Somalia.

Participants: Randomly sampled individuals of all ages.

Main outcome measure: Cartographic descriptions of malaria receptivity and contemporary risks in Somalia at the district level.

Results: The contemporary annual $PfPR_{2-10}$ map estimated that all districts (n=74) and population (n=8.4 million) in Somalia were under hypoendemic transmission (≤ 10% $PfPR_{2-10}$). Of these, 23% of the districts, home to 13% of the population, were under transmission of <1% $PfPR_{2-10}$. About 58% of the districts and 55% of the population were in the risk class of 1% to <5% $PfPR_{2-10}$. In contrast the receptivity map estimated 65% of the districts and 69% of the population were under mesoendemic transmission (>10% to 50% $PfPR_{2-10}$) and the rest as hypoendemic.

Conclusion: Compared to maps of receptive risks, contemporary maps of transmission mask disparities of malaria risk necessary to prioritize and sustain future control. As malaria risk declines across Africa, efforts must be invested in measuring receptivity for efficient control planning.

Article focus

- Cross-sectional *PfPR* prevalence survey data for the period 2007 to 2010 in Somalia
- Bayesian geostatistical models estimating the receptive and contemporary malaria transmission in Somalia
- Implications of the two malaria risk maps for malaria control planning in Somalia

Key messages

- It is feasible to use *Pf*PR community prevalence data in space and time to estimate the receptive and contemporary risks of malaria within the same model framework.
- Malaria receptivity maps are critical to inform the scale-up and sustaining of interventions where disease has declined or is highly seasonal.
- Efforts must be invested in helping malaria endemic countries in Africa measure their receptive risks.

Strengths and limitations

- The annual *Pf*PR surveys provide unique opportunities to measure receptivity in Somalia.
- Improving the spatial and temporal distributions of *Pf*PR data and exploring probabilistic approaches of selecting potential maximum risks will improve the measurement of receptivity.

INTRODUCTION

Malaria receptivity is a measure of the intrinsic vector transmission potential of an area[1]. Interest in measuring malaria receptivity has emerged following the resurgence of the malaria elimination agenda[2,3] and the need to quantify the risks posed by human population movement leading to the reintroduction of transmission[4,5]. However, understanding receptivity is equally important to decision-making for countries that are implementing control. In low stable endemic countries national programmes need to understand the risks posed by withdrawal of interventions from areas that are historically high transmission[3,6]. In unstable transmission areas where parasite exposure is highly seasonal and prone to climatic anomalies, targeting interventions to prevent the risk of epidemics are a priority[1]. The empirical malaria risk maps that are commonly available to countries to support malaria control planning are those that represent the contemporary distribution of risk under control[7,8] and are therefore of limited value in defining the epidemic potential or the receptive rebound risks of withdrawing or failing to sustain interventions.

Measuring receptivity for malaria control countries ideally requires empirical data from a period of no control and under the optimum transmission conditions. There is hardly any country in Africa that has remained universally control-naive over the last 100 years[9]. Alternatively, nationally representative empirical data on malaria transmission during the pre-Roll Back Malaria (RBM) era, or before intervention scale-up reached critical thresholds for a given country, may represent the best approximation of receptivity. Furthermore, such information must be resolved to administrative decision-making units to which malaria resources are allocated to make them relevant for policy. In this study we use community *Plasmodium falciparum* parasite prevalence data from 2007 to 2010 within a model-based geostatistical (MBG) framework to develop contemporary and receptive risk maps and resolve endemicity to districts in Somalia.

METHODS

Country context

Somalia is divided into the three zones of North West (Somaliland), North East (Puntland) and Central South (Figure 1). The northern zones are generally dry and hot whereas the Central South zone has sub-tropical climate and is where the two major rivers of the country, the Shabelle and the Juba, are located[10-11]. *Anopheles arabiensis* is the dominant malaria transmitting vector throughout the country although *An. funestus* is reported in Central South[12, 13]. *P. falciparum* is the dominant species of the malaria parasite[14-16]. The presence of *P. vivax* cases have also been reported with studies in Somaliland showing relatively high vivax antibody responses[17]. The failure of the long rains in Somalia in 2010 combined with the below average rainfalls in previous two seasons have resulted in a severe drought[11].

Since 1991 Somalia has had no effective central government and has experienced frequent internal armed conflicts resulting in the breakdown of public services and political disintegration[18]. Although the transitional federal government (TFG), formed in February 2004, is internationally recognized as the official government of Somalia[19], in reality the three zones currently function as semi-independent states. Consequently, there are three ministries of health but the majority of health care is provided by international and national non-government organizations that have come together under the umbrella of Somalia Aid Coordinating Board which was later reformed as the Somalia Support Secretariat[20].

The main funder of malaria control in Somalia is the Global Fund to Fight Aids TB and Malaria (GFATM) with the United Nations Children Fund (UNICEF) as the principal recipient and the World Health Organization (WHO) as sub-recipient[21,22]. In 2010, the second national malaria strategy was launched with universal scale-up of vector control, parasitological diagnosis, effective antimalarials, improved surveillance and epidemic preparedness and response as the main strategic approaches[23]. Since 2004, over 77 million USD has been approved to Somalia by the GFATM for malaria control resulting in the distribution of almost one million LLINs[24].

Community survey data

The community *P. falciparum* parasite rate (*Pf*PR) is the most commonly used indicator for mapping malaria transmission[7]. This is because it is easy to measure, has a historical legacy and a predictable relationship with other measures of transmission intensity such as entomological inoculation rate and the basic reproductive rate[25,26]. The *Pf*PR data used for the present study were assembled through the Food and Agriculture Organization-Food Security and Nutrition Analysis Unit (FAO-FSNAU) surveys undertaken regularly in

Somalia[16,27]. These surveys were initially established to monitor the nutritional status of children less than 5 years of age and began in 2000[27]. Investigations of malaria prevalence covering persons of all ages were only included from 2007[16] and have been undertaken annually since, covering most regions of Somalia. A detailed description of the sampling design is provided elsewhere[16]. During the survey, respondents provided a finger prick blood sample that was examined for the presence of *P. falciparum* infection using a rapid diagnostic test (RDT) (Paracheck Pf TM, Orchid Biomedical Systems, Goa, India). Consent was obtained for all individuals before interview and separately for the malaria testing. Additional information was recorded on the date of survey and age and sex of participants. After all survey data were assembled each surveyed community was geo-coded using combinations of global positioning systems (GPS), electronic gazetteers (Google Earth, Encarta and Alexandria), and other sources of longitude and latitude such a settlement database collated by FAO-SWALIM[11].

Assessment of ecological and climatic predictors of malaria risk

The transmission intensity of malaria is influenced by climatic and ecological factors through their independent or combined effects on the survival of the Anopheline vectors and the Plasmodium parasites within the vector[28]. The ecological and climatic factors used commonly to improve the precision of empirical malaria mapping are urbanisation, rainfall, temperature and distance to potential mosquito larva breeding sites or their proxies.

Data at 1 × 1 km spatial resolution on urbanisation[19,30], annual mean precipitation[31,32]and enhanced vegetation index (EVI)[33] were assembled. Maps of rivers, floodplains, reservoirs and coastal wetlands were assembled from the Global Wetlands and Lakes Database[34] and Euclidean distances to these proximate of breeding sites were computed in ArcGIS 10 (ESRI Inc. NY, USA). As a metric for the effect of temperature on malaria transmission, a temperature suitability index (TSI) at a spatial resolution of 1 × 1 km was used. TSI was constructed using monthly temperature time series within a biological modelling framework to quantify the effect of ambient temperature on sporogony and vector survivorship and determine the suitability of an area to support transmission globally[35]. The values of the underlying ecological and climatic covariates were extracted to each survey location using ArcGIS 10 *Spatial Analyst* tool. Distance to potential breeding sites was log-transformed before analysis because of its high positive skew. The covariates were then included in total-sets analysis which is an automatic model selection process based on a generalized linear regression model and implemented using the *bestglm* package in R[36,37]. This approach selects the best combination of the covariates based on the value of the Bayesian Information Criteria (BIC) statistic[38] which selects the lowest BIC

as the best model fit. Details of the ecological and climatic predictors of *Pf*PR and the results of the total-set analysis are provided in Supplementary Information (SI) 1.

The space-time Bayesian geostatistical model for predicting *P. falciparum* distribution in Somalia

Space-time MBG methods offer the flexibility of predicting an outcome to any given year in a time series by harnessing fully both the spatial and temporal relationships of the data and generate uncertainties of the predictions from the full posterior distributions[39].

In this study, the assembled *Pf*PR data were standardized to the classical age-range of 2 to less than 10 years using an algorithm based on modified catalytic conversion models[40]. The continuous surfaces of the age-standardised data ($PfPR_{2\text{-}10}$) were generated using a space-time MBG framework[8,41] whereby Bayesian inference was implemented using the Markov Chain Monte Carlo algorithm. Details of model code[41] and statistical procedures[8] are provided in SI 2. In brief, the value of $PfPR_{2\text{-}10}$ was modelled as a transformation of a spatiotemporally structured field superimposed with unstructured (random) variation on a regular 1 × 1 km grid from 2007 to 2010. The number of *P. falciparum* positive responses from the total sample at each survey location was modelled as a conditionally independent binomial variate given the unobserved underlying age-standardised $PfPR_{2\text{-}10}$ value [40] [Smith et al 2007] and a linear function of the climatic and environmental predictors. The unstructured component was represented as Gaussian distribution with zero mean. The spatiotemporal component was represented by a stationary Gaussian process[42] with covariance defined by a spatially anisotropic version of the space-time covariance function proposed by Stein (2005)[43]. To partly model seasonality, the covariance function was modified to allow the time-marginal model to include a periodic component of wavelength 12 months in the temporal covariance structure. Each survey was referenced temporally using the mid-point (in decimal years) between the recorded start and end months. For each grid location samples of the annual mean of the full posterior distribution of $PfPR_{2\text{-}10}$ for each year were generated. These $PfPR_{2\text{-}10}$ samples were then used to generate continuous maps of the annual mean. To determine the probable maximal malaria risk the highest value of predicted mean annual $PfPR_{2\text{-}10}$ value at each 1 × 1 km grid location over the period 2007 to 2010 were extracted. These were then combined to generate a single map of maximum mean $PfPR_{2\text{-}10}$.

Assessing uncertainty of model predictions

As a first step to understanding the uncertainty around the predictions of $PfPR_{2\text{-}10}$ using the Bayesian geostatistical model, the continuous mean maps were accompanied by estimates of the posterior standard deviation. For the maximum mean $PfPR_{2\text{-}10}$ map, the posterior standard deviations associated with the selected

mean value was used. To allow for a scaled comparison of the uncertainty of the 2010 *Pf*PR_{2-10} map and the 2010 annual mean *Pf*PR_{2-10}, the coefficient of variation, which is a measure of dispersal around the mean[44], was computed as the ratio of the standard deviation to the mean. Higher values of the coefficient of variation suggest increasing uncertainty of model predictions. In addition, a spatially representative validation set of *Pf*PR_{2-10} survey data were also selected using a spatially declustered sampling algorithm[41]. The annual predictions were then repeated in full using the remaining data to predict mean *Pf*PR_{2-10} at the validation locations. The ability of the model to predict point-values of *Pf*PR at unsampled locations was quantified using two simple summary statistics: the mean prediction error (ME); and the mean absolute prediction error (MAE). The ME provides a measure of the model bias, while the MAE is a measure of the average accuracy of individual predictions.

Defining district level malaria endemicity and population at risk

A 2010 population surface for Somalia at 100 × 100 m spatial resolution was provided by the AfriPop project[29,30]. Using this map, estimates of the total population of each 1 × 1 km pixel to which mean *Pf*PR_{2-10} was predicted was computed in ArcGIS 10. The 1 × 1 km grid squares with attached population estimates were further classified separately by the mean *Pf*PR_{2-10} 2010 and the maximum mean *Pf*PR_{2-10}. To weight endemicity for population distribution, only those grid squares with population were retained and from these the mean *Pf*PR_{2-10} was computed for each district. Based on the aggregate mean *Pf*PR_{2-10} districts were then classified using a modification of the classical malaria endemicity classification[45]. The hypoendemic class (≤10% *Pf*PR_{2-10}) was split into <1% *Pf*PR_{2-10}; 1% to <5% *Pf*PR_{2-10} and 5% to 10% *Pf*PR_{2-10}. The endemicity class of <1% *Pf*PR_{2-10} represents the threshold at which an area is considered to be under low stable endemic control and a decision for sustaining control or aiming for elimination can be made[3,6]. The hyperendemic (>50% to 75% *Pf*PR_{2-10}) and holoendemic (>75% *Pf*PR_{2-10}) classes were also combined while the mesoendemic class (>10% to 50% *Pf*PR_{2-10}) remained unchanged. The number of districts and population by these endemicity classes were then summarised based on both the 2010 mean *Pf*PR_{2-10} (contemporary risk) map and the maximum mean *Pf*PR_{2-10} (receptive risk) map.

RESULTS

Predictions of mean annual $PfPR_{2-10}$ to 2010 and maximum mean posterior $PfPR_{2-10}$

A total of 1,558 *P. falciparum* community surveys (Figure 1) in which 103,593 persons were examined were assembled for the period 2007-2011 in Somalia. The majority of the data were located in the Central South zone where most of the population live. Survey data were collected across nine different months over the four years with the majority of data (76%) assembled in the months of November, December, May and June

corresponding to the peak malaria seasons in Somalia. The results of the total-set analysis showed that the model with urbanisation, precipitation, EVI and distance to main water bodies and floodplains as the best fit in predicting *Pf*PR and these variables were subsequently included in the malaria prediction model (SI 1). TSI was not selected as a statistically strong predictor of *P. falciparum* prevalence in Somalia.

The continuous 2010 malaria endemicity map for Somalia *Pf*PR_{2-10} showed the majority of locations were predicted to have parasite prevalence of <5% indicating largely hypoendemic transmission (Figure 2A). The majority of areas in North East and North West zones were predicted to be under *Pf*PR_{2-10} <2%. In contrast, the maximum annual mean *Pf*PR_{2-10} map showed a substantially different risk landscape with the majority of the Central South zone having *Pf*PR_{2-10} of >10% and an a maximum predicted mean of 38%, suggesting that peak malaria transmission in all of Central South zone and southern parts of North East zone is mesoendemic (Figure 2B). In the northern zones, maximum mean risks were predicted to be predominantly between 4% to <10% *Pf*PR_{2-10}.

The ME and MAE associated with the full space-time geostatistical model was 4.8% and 0.2% respectively. The 2010 annual mean *Pf*PR_{2-10} predictions were associated with higher coefficients of variation compared to the maximum mean *Pf*PR_{2-10} predictions although the difference was moderate (Figure 2C and Figure 2D). In both maps, uncertainty appeared highest in northern zones where data in space and time were fewest.

District estimates of contemporary and receptive malaria risk

According to the contemporary district malaria endemicity map based on the annual mean *Pf*PR_{2-10} map of 2010 all districts in Somalia were under hypoendemic transmission (Figure 3A). Out of the 74 districts an estimated 17 (23%) districts covering about 1.1 million people (13%) were in the <1% *Pf*PR_{2-10} risk class (Table 1 and Figure 3A). The majority of the districts (58%) and population (55%) were in the risk class of 1% to <5% *Pf*PR_{2-10} and the rest were under risks of 5% to 10% *Pf*PR_{2-10}.

In contrast the receptive risk map showed that there were no district under low stable endemic control (<1% *Pf*PR_{2-10}) and the majority of the districts (65%) and population (69%) were under the mesoendemic class (Table 1 and Figure 3B) with an upper district maximum mean estimate of 35% *Pf*PR_{2-10}. About 27% of the districts and 17% of the population were in the upper range of hypoendemicity (5% to 10% *Pf*PR_{2-10}). The rest of the districts and population were in the intermediate hypoendemic class of 1% to <5% *Pf*PR_{2-10}.

DISCUSSION

The malaria risk maps that are commonly available to countries in Africa to support malaria control planning are those that represent the contemporary distribution of risk[8, 16, 46-50]. They have been developed primarily from geo-coded parasite rate survey data[7] usually to predict risk to the most recent data year and therefore reflect transmission under scaled interventions during the era of the RBM partnership[51, 52]. In this study we argue that, in addition to contemporary maps of malaria risks, low stable endemic control and unstable transmission countries require maps of receptivity to assess the risks of rebound and epidemics and decide on where to scale-up and/or sustain intervention coverage. To demonstrate this we used community *Pf*PR survey data from the period 2007 to 2010 within a space-time MBG framework to generate two continuous malaria risk maps for Somalia. One is a contemporary map of annual mean $PfPR_{2-10}$ predicted to 2010 (Figure 2A) and the other is the maximum annual mean $PfPR_{2-10}$ map derived from the highest mean $PfPR_{2-10}$ value predicted to a location in any year over the study period to approximate receptivity (Figure 2B). We resolved these maps to the district, which is the malaria decision making unit in Somalia, and classified them by endemicity using population-weighted mean $PfPR_{2-10}$ (Figure 3).

The efficacy and impact of malaria interventions on disease in an area are dependent on its intrinsic transmission potential[53-55]. This is the theoretical basis upon which international guidelines for malaria control are formulated[1,56]. One of the most important applications of malaria risk maps for control planning, therefore, is to inform the spatial targeting of the appropriate mixes of interventions[7,57]. For Somalia, the results of the comparison of the contemporary and the receptive risk maps represents two very different transmission scenarios (Table 1; Figure 3). The contemporary malaria risk map predicted that all of Somalia was under conditions of hypoendemic transmission (≤10% $PfPR_{2-10}$) in 2010 with a fifth of the districts under risks of <1% $PfPR_{2-10}$ while the majority of the districts and population were in the intermediate hypoendemic transmission class of 1% to <5% $PfPR_{2-10}$. Under these transmission conditions targeted distribution of long lasting insecticidal nets (LLINs) and indoor residual spraying (IRS) aimed at control of residual foci are recommended while intermittent presumptive treatment in pregnancy (IPTp) is not[56]. Instead of being part of the broader monitoring and evaluation process, disease surveillance is also regarded as an intervention in of itself[1] comprising high quality passive case detection, case notification and active case detection in which all febrile cases within proximity of the index case are tested and those positive for malaria infection are radically treated[58]. In the districts where transmission is <1% $PfPR_{2-10}$, malaria elimination is considered to be technically feasible[3,6] presenting an opportunity to re-orient the sub-national strategy here towards elimination and undertake an assessment of its operational feasibility[5]. In contrast, the receptive risk map predicted that over 65% of the districts and population were under mesoendemic transmission (>10% to 50% $PfPR_{2-10}$) with the

rest exposed to hypoendemic transmission. Using this map, in the hypoendemic districts LLINs and IRS would be better targeted to foci of risk and in preparation for possible epidemics as universal coverage with these interventions is unlikely to be the most cost-effective. In those areas of receptive mesoendemic transmission, which comprise 65% of the districts, universal coverage with LLINs should be the sustained ambition[1,2,56]. The two divergent potential national malaria strategies emanating from the two different descriptions of risk highlight the danger of relying on contemporary risk maps to make decisions that require the understanding of the intrinsic transmission potential of malaria.

Available maps that describe pre-RBM distribution of risks are either expert opinion maps[59] or climate-based deterministic transmission suitability models[60] and not driven by empirical data. Even where empirical data may be available, in countries with unstable malaria transmission susceptible to seasonal and anomalous climatic variations such as the recent drought in Somalia, the risks measured at one time point may not be representative of the possible peak risk levels for that point. Therefore spatially nationally representative data over several years are required to capture these variations and estimate the highest possible transmission. In this study, *Pf*PR data for Somalia that is available over four consecutive years has provided a unique opportunity to develop a novel way of selecting the maximum predicted risks within the time series. The resolution of risk levels at the malaria resource decision-making unit also represents a product that is likely to be of more policy relevance to the national programme managers compared to the more common pixel-level predictions of risks.

The study has some limitations. Although Somalia represents one of the few African countries with ubiquitous *Pf*PR data in space and time there are gaps in the distribution of the data and uncertainty of the predictions are partly a function of these. The validation tests, however, show overall good predictive model performance with overall bias of ME of <5% and a slight average over-prediction of about MAE of 0.2%. The coefficient of variation, which is the ratio of standard deviation to mean $PfPR_{2\text{-}10}$, appeared similar for both the 2010 mean and the maximum mean maps with uncertainty highest in northern zones where data in space and time were fewest (Figure 2 C & D). In selecting the maximal mean risk as predicted to a location in any year over the four year period, we make assumptions as if the modelled predictions were part of four-year repeat 'observations' of $PfPR_{2\text{-}10}$, in that location. The basis for this is that if the mean estimate at any 1 × 1 km location from the full posterior distribution of the space-time model is a robust estimate to the given time and location, then selected maximum mean estimate is equally so. We suggest that this is a plausible assumption but further efforts need to be invested in probabilistically selecting the maximum mean predictions from the series of usually spatially and temporally uneven data. Any uncertainties in the approach used to select the maximum

annual mean *Pf*PR$_{2-10}$ are however unlikely to be the source of the major differences in endemicities when compared to the annual mean *Pf*PR$_{2-10}$ for 2010.

To compute a single estimate of risk for a relatively large area, such as districts in Somalia, will always obscure some of the heterogeneity in malaria distribution within that area regardless of the methods used. Any decision to do so is therefore a compromise between the practical applications of such a classification and the potential loss of precision in risk estimation. Approaches that directly adopt the heterogeneous properties of the prevalence data to make statistically robust single estimates of mean *Pf*PR to an administrative unit are computationally and methodologically intensive[61] but have the advantage of estimating the area level uncertainty classification through joint simulation. In this study we have used simpler approaches to partly capture the within-district heterogeneity in malaria risk when classifying them into a single endemicity class by first assigning pixel-level population to an endemicity class before aggregating to the district. Future efforts should explore approaches such as joint simulation[61] and small area estimation[62] techniques to describe the uncertainties around area level estimates of risks robustly. Such measures of uncertainties are not only quantitative estimates of model validity but also help determine where future data assembly must be prioritised to improve precision.

In conclusion, the aim of this study was to demonstrate the need for malaria receptivity maps for optimal malaria resource planning in countries which have either achieved low stable endemic control or are of unstable transmission and therefore susceptible to seasonality of climatic anomalies. We have used approaches that derive maps of contemporary malaria risk and approximations of receptivity within the same space-time MBG model resolved at the district level in Somalia. The two maps show significantly divergent transmission scenarios in which the contemporary map describes the majority of Somalia as hypoendemic and while the other shows a largely mesoendemic transmission profile. These disparities have far-reaching consequences on decisions regarding the design and scale-up of interventions in Somalia. The results have important control implications for several low transmission countries in Africa. Urgent efforts must therefore be invested in assembling detailed historical data on parasite prevalence to allow for a better understanding of receptivity and equip national programmes with reliable estimates of receptivity that will enhance better decision making.

Acknowledgements The community survey parasite prevalence data was provided by the Food and Agricultural Organization-Food Security and Nutritional Analysis Unit (FAO-FSNAU) for Somalia. The authors are grateful to the support and diligence of the all the FSNAU field staff who did a great job under difficult conditions. We are grateful to all household members who participated in the survey and agreed to the malaria testing. We thank Jacob Ouko and Damaris Kinyoki for their help with assembly of ancillary data. The authors are also grateful for comments on earlier drafts of the manuscript from Dr Emelda Okiro. This paper is published with permission from the Director of KEMRI.

Contributors AMN was responsible for overall scientific management, study design, data cleaning, analysis, interpretation, drafting and production of the final manuscript. VAA was responsible for data cleaning, geo-coding, analysis and contributed to the final manuscript. APP was responsible for the coding of the MBG models, developed the supplementary information on model specifications and contributed to final manuscript. GM and MB contributed to the survey design, data assembly and cleaning and contributed to final manuscript. FY and JA contributed to survey design, interpretation of results and contributed to final manuscript. RWS provided scientific guidance and contributed to the analysis, interpretation and preparation of the final manuscript. All authors read and approved the final manuscript.

Funding: Cross-sectional survey was funded by the FAO-FSNAU and partners. AMN is supported by the Wellcome Trust as an Intermediate Research Fellow (#095127). RWS is supported by the Wellcome Trust as Principal Research Fellow (#079080) that also funded support to APP. Programmatic support for this study was also provided through a Wellcome Trust Major Overseas Programme grant to the KEMRI/Wellcome Trust Research Programme (#092654).

Competing Interests None**.**

Ethics approval Ethical approval was provided through permission by the Ministry of Health Somalia, Transitional Federal Government of Somalia Republic, Ref: MOH/WC/XA/146./07, dated 02/02/07. Informed verbal consent was sought from all participating households and individuals. Participants who were positive for *Plasmodium falciparum* infection was treated with the correct dosages of the nationally recommended antimalarial.

Data sharing statement Data from this study are not in the public domain

References

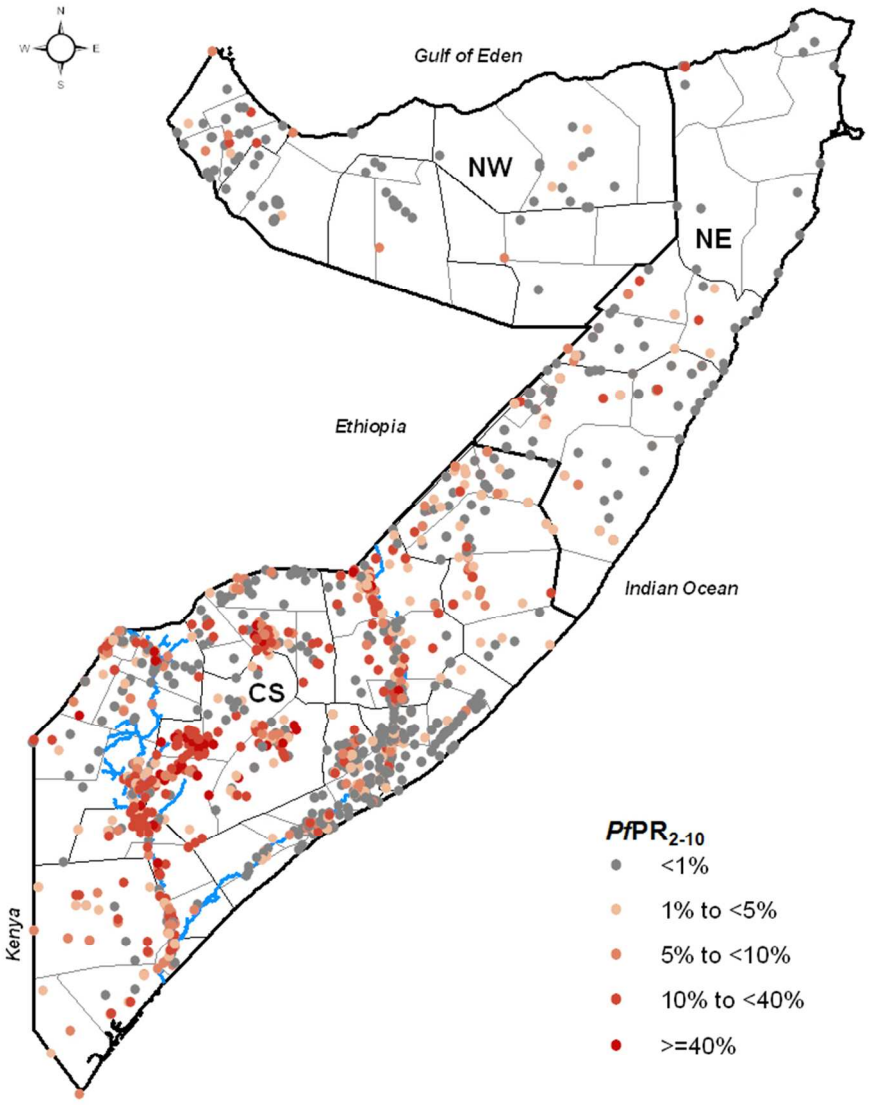
1. World Health Organization: Malaria elimination: a field manual for low and moderate endemic countries, Geneva: World Health Organization; 2007.
2. Feachem RGA, Phillips AA, Hwang J, Cotter C, Wielgosz B, Greenwood BM, *et al*. Shrinking the malaria map: progress and prospects. *Lancet* 2010; 376: 1566 – 1578.
3. Cohen JM, Moonen B, Snow RW, Smith DL. How absolute is zero? An evaluation of historical and current definitions of malaria elimination. *Malar J* 2010; 9: 213.
4. Tatem AJ, Qiu Y, Smith DL, Sabot O, Ali AS, Moonen B. The use of mobile phone data for the estimation of the travel patterns and imported Plasmodium falciparum rates among Zanzibar residents. *Malaria J* 2009; 8: 287.
5. Tatem AJ, Smith DL. International population movements and regional *Plasmodium falciparum* malaria elimination strategies. *Proceedings of the National Academy of Sciences of the United States of America*, 2010; 107: 12222 - 12227.
6. Snow RW, Marsh K. Malaria in Africa: progress and prospects in the decade since the Abuja Declaration. *Lancet* 2010; 376: 137-139.
7. Hay SI, Snow RW. The Malaria Atlas Project: Developing global maps of malaria risk. *PLoS Med* 2006; 3: e473.
8. Gething PW, Patil AP, Smith DL, Guerra CA, Elyazar IRF, Johnston GL, *et al*. A new world malaria map: Plasmodium falciparum endemicity in 2010. *Malar J* 2011;10: 378 doi: 10.1186/1475-2875-10-378.
9. Snow RW, Amratia P, Kabaria CW, Noor AM. The changing limits and incidence of malaria in Africa: 1939-2009. *Adv Parasitol* 2012; in press.
10. Hadden RL. *The Geology of Somalia: A Selected Bibliography of Somalian Geology, Geography and Earth Science*. Engineer Research and Development Laboratories, Topographic Engineering Centre. http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA464006&Location=U2&doc=GetTRDoc.pdf. Accessed 12 December 2011.
11. FAO-SWALIM Databases. http://geonetwork.faoswalim.org/geonetwork/srv/en/main.home. Accessed 10th December 2011.
12. Choumara R. Notes sur le paludisme au Somaliland. *Riv di Malariol* 1961; 40: 9-34.
13. Kamal M. *Entomological surveillance in Somalia*. In Consultancy report for WHO Somalia, 2007.
14. Ilardi I, Sebastian A, Leone F, Madera A, Khalif BM, Shukri CS, *et al*. Epidemiological study of parasitic infections in Somali nomads. *Trans R Soc Trop Med Hyg* 1987; 81: 771-772.
15. Warsame M, Perlmann H, Ali S, Hagi H, Farah S, Lebbad M, *et al.* The sero-reactivity against Pf155 (RESA) antigen in villagers from a meso-endemic area in Somalia. *Trop Med & Parasitol* 1989; 40: 412-414.
16. Noor AM, Clements ACA, Gething PW, Moloney G, Borle M, Shewchuk T, *et al.* Spatial prediction of *Plasmodium falciparum* prevalence in Somalia. *Malar J* 2008; 7: 159.
17. Bousema T, Jackie C, Patrick C, Jonathan C, Immo K, Eleanor R, *et al*. Serological markers for low exposure to malaria. *Emerg Infect Dis* 2009; 16: 392-399.
18. The Crisis Group. Somalia. http://www.crisisgroup.org/en/regions/africa/horn-of-africa/somalia.aspx, Accessed 12 November 2011.
19. The Transitional Federal Government of Somalia http://en.wikipedia.org/wiki/History_of_the_Transitional_Federal_Government_of_the_Republic_of_Somalia, Accessed 12 December 2011.
20. UNOPS 2010. The Somali Support Secretariat Project Board Terms of Reference as promulgated at the 10 May 2010 Project Board Meeting. http://www.coordinate4somalis.info/images/stories/pdf/Somali%20Support%20Secretariat%20Project%20Board%20Terms%20of%20Reference.pdf. Accessed 7 March 2012.
21. GFATM Project proposal.http://www.theglobalfund.org/search/docs/6SOMM_1418_0_full.pdf. Accessed 10 December 2007.
22. World Health Organization-Roll Back Malaria. Somalia National Strategic Plan for Malaria 2011 – 2015; 2010.
23. World Health Organization-Roll Back Malaria. National Malaria Prevention and Control Monitoring and Evaluation Plan 2011 – 2015; 2010.
24. GFATM Progress report. http://www.theglobalfund.org/en/library/publications/progressreports/. Accessed 15 December 2011.
25. Smith DL, McKenzie FE, Snow RW, Hay SI. Revisiting the basic reproductive number for malaria and its implications for malaria control. *PLoS Biol* 2007; 5: e42.

26. Smith DL, Dushoff J, Snow RW, Hay SI. The entomological inoculation rate and *Plasmodium falciparum* infection in African children. *Nature* 2005; 438: 492-495.
27. FSANU Analytical systems. http://www.fsausomali.org/200510112504_baselines.php. Accessed 11th January 2012.
28. Molineaux L, Muir DA, Spencer HC, Wernsdorfer WH. *The epidemiology of malaria and its measurement*. Malaria: Principles and Practice of Malariology. W. H. Wernsdorfer and I. McGregor. London, Churchill Livingstone, 1988; 2: 999-1089
29. Linard C, Alegana VA, Noor AM, Snow RW, Tatem AJ. A high resolution spatial population database of Somalia for disease burden estimation. *Intl J of Health Geogr* 2010; 9: 45.
30. The AfriPop Project. http://www.clas.ufl.edu/users/atatem/index_files/Details.htm, Accessed 15 January 2011.
31. Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A.Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol* 2005; 25: 1965-1978.
32. WorldClim-Global Climate Database. www.wordlclim.org. Accessed 10 February 2012.
33. MODIS- EVI data archives. ftp://n4ftl01u.ecs.nasa.gov/SAN/MOST/MOD10A2.005/. Accessed 12 September 2011.
34. Global Lakes and Wetlands Database Request. https://secure.worldwildlife.org/science/data/item1877.html. Accessed 10 September 2011.
35. Gething PW, Van Boeckel T, Smith DL, Guerra CA, Patil AP, Hay SI, *et al*. Modelling the global constraints of temperature on transmission of *Plasmodium falciparum* and *P. vivax*. *Parasite Vector* 2011; 4:92.
36. Miller A. Subset Selection in Regression. Boca Raton, FL: Chapman & Hall. 2002:238 p.
37. Lumley T leaps: regression subset selection (R package) Version 2.7, 2010.
38. Schwarz G. Estimating dimensions of a model. *Ann Stat* 1978; 6: 461-464.
39. Diggle P, Moyeed R, Rowlingson B, Thompson M. Childhood malaria in the Gambia: a case-study in model-based geostatistics. *Appl Stat* 2002; 51: 493-506.
40. Smith DL, Guerra CA, Snow RW, Hay SI. Standardizing estimates of the Plasmodium falciparum parasite rate. *Malar J* 2007; 6: 131.
41. Malaria Atlas Project *P. falciparum* Cartographic code. https://github.com/malaria-atlas-project/mbgw-clean. Accessed October 2011.
42. Banerjee S, Carlin BP, Gelfand AE. *Hierarchical modeling and analysis for spatial data. Monographs on Statistics and Applied Probability 101*. Boca Raton, Florida, U.S.A.: Chapman & Hall / CRC Press LLC, 2004.
43. Stein ML. Space-time covariance functions. *J Am Stat Assoc* 2005; 100: 310-321.
44. Kirkwood TBL. Geometric means and measures of dispersion. *Biometrics* 1979; 35:908-909.
45. Metselaar D, PH van Thiel. Classification of malaria. *Trop Geogr Med* 1959; 11: 157-161.
46. Noor AM, Gething PW, Alegana VA, Patil AP, Hay SI, Muchiri E, *et al*. The risks of malaria infection in Kenya in BMC *Infect Dis* 2009; 9*:* e180.
47. Gosoniu L, Veta AM, Vounatsou P. Bayesian Geostatistical Modeling of Malaria Indicator Survey Data in Angola. *PLoS ONE* 2010*;* 5: e9322. doi:10.1371/journal.pone.0009322.
48. Riedel N, Vounatsou P, Miller JM, Gosoniu L, Chizema-Kawesha E, Mukonka V, *et al*. Geographical patterns and predictors of malaria risk in Zambia: Bayesian geostatistical modelling of the 2006 Zambia national malaria indicator survey (ZMIS). *Malar J* 2010; 9: 37.
49. Taylor SM, Messina JP, Hand CC, Juliano JJ, Muwonga J, Tshefu AK, *et al*. Molecular Malaria Epidemiology: Mapping and Burden Estimates for the Democratic Republic of the Congo, 2007. *PLoS ONE* 2011; 6(1): e16420. doi:10.1371/journal.pone.0016420
50. Stensgaard A-S, Vounatsou P, Onapa AW, Simonsen PE, Pedersen EM, Rahbek C, *et al*. Bayesian geostatistical modelling of malaria and lymphatic filariasis infections in Uganda: predictors of risk and geographical patterns of co-endemicity. *Malar J* 2011; 10:298
51. Roll Back Malaria. Progress and Impact Series Number 2. http://www.rbm.who.int/ProgressImpactSeries/docs/RBMMalariaFinancingReport-en.pdf, Accessed 16 February 2012.
52. Roll Back Malaria. Progress and Impact Series Number 2. http://www.rbm.who.int/ProgressImpactSeries/docs/wmd2010report-en.pdf, Accessed 16 February 2012.
53. Snow RW, Marsh K. The consequences of reducing transmission of Plasmodium falciparum in Africa. *Adv Parasitol* 2002; 52: 235-264.
54. Killeen GF, Smith TA, Ferguson HM, Mshinda H, Abdulla S, Lengeler C, *et al.* Preventing Childhood Malaria in Africa by Protecting Adults from Mosquitoes with Insecticide-Treated Nets. *PLoS Med* 2007; 4: e229.

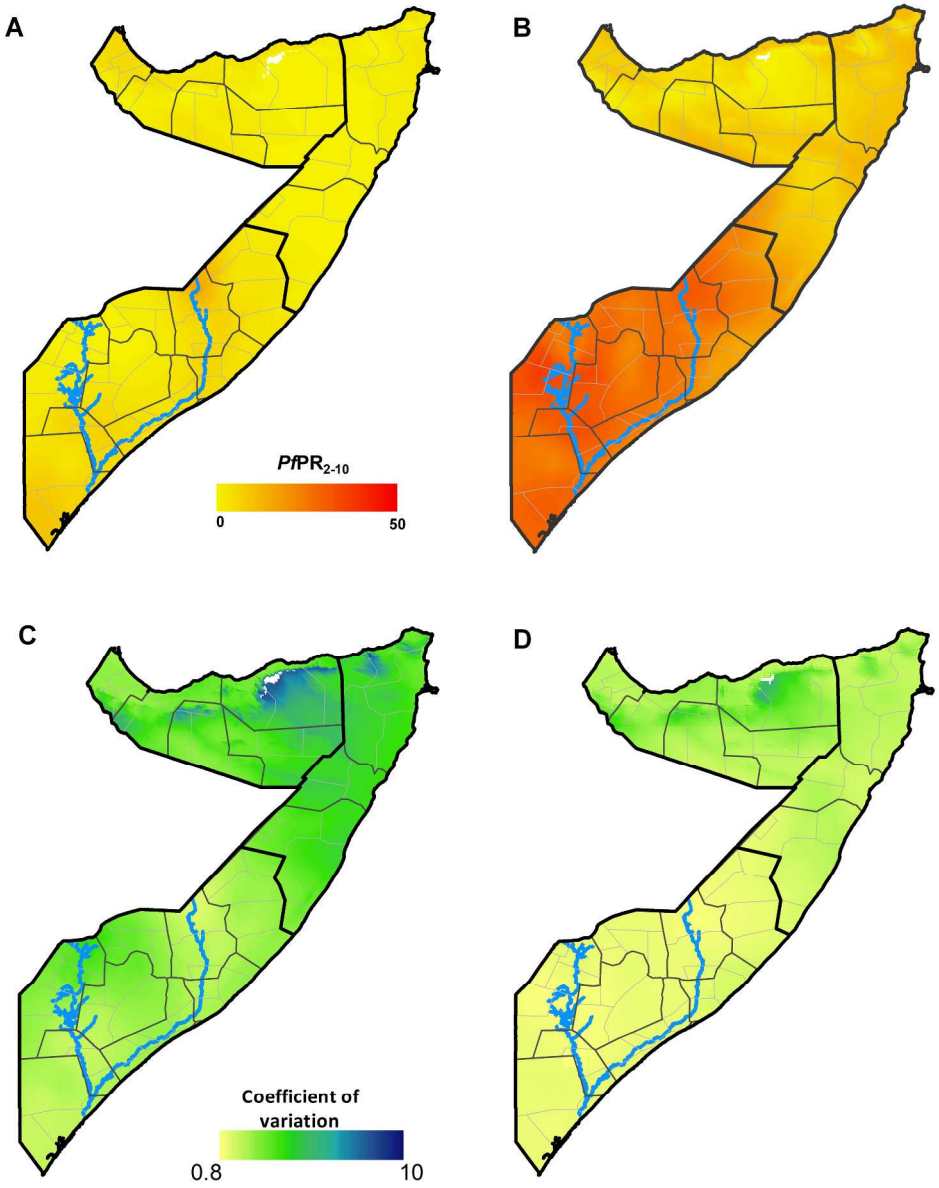
55. Smith DL, Noor AM, Hay SI, Snow RW. Predicting changing malaria risk following expanded insecticide treated net coverage in Africa. *Trends Parasitol* 2009, 25: 511-516.
56. World Health Organization. Manual for developing a national malaria strategic plan, Geneva, WHO, September 2011.
57. Noor AM, Alegana VA, Patil AP, Snow RW. Predicting the unmet need for biologically targeted coverage of insecticide treated nets in Kenya. *Am J Trop Med Hyg* 2010*,* 83: 854-860.
58. Moonen B, Cohen JM, Snow RW, Slutsker L, Drakeley C, Smith DL, *et al*. Operational strategies to achieve and maintain malaria elimination. *Lancet* 2010; 376: 1592 – 1603.
59. Lysenko AJ, Semashko IN: Geography of malaria. A medico-geographic profile of an ancient disease [in Russian]. In: *Itogi Nauki: Medicinskaja Geografija.* Edited by Lebedew AW.Moscow: *Acad. Sci*, USSR; 1968: 25-146.
60. Craig MH, Snow RW, le Sueur D. A climate-based distribution model of malaria transmission in sub-Saharan Africa. *Parasitol Today* 1999; 15: 105–111.
61. Gething PW, Patil AP, Hay SI. Quantifying aggregated uncertainty in *Plasmodium falciparum* malaria prevalence and populations at risk via efficient space-time geostatistical joint simulation. *PLoS Comput Biol* 2010*;* 6: e1000724.
62. Bayesian methods for combining multiple Individual and Aggregate data Sources in observational studies. http://www.bias-project.org.uk/software/#sae. Accessed 15 February 2011.

Table 1 A summary of districts (N=74) and population in 2010 (N=8.4 million) in Somalia classified by malaria endemicity. District classifications of endemicity were computed from population-weighted posterior annual mean *Pf*PR_{2-10} predicted to 2010 (contemporary) and the maximum annual mean *Pf*PR_{2-10} (receptive) predictions over the period 2007-2010.

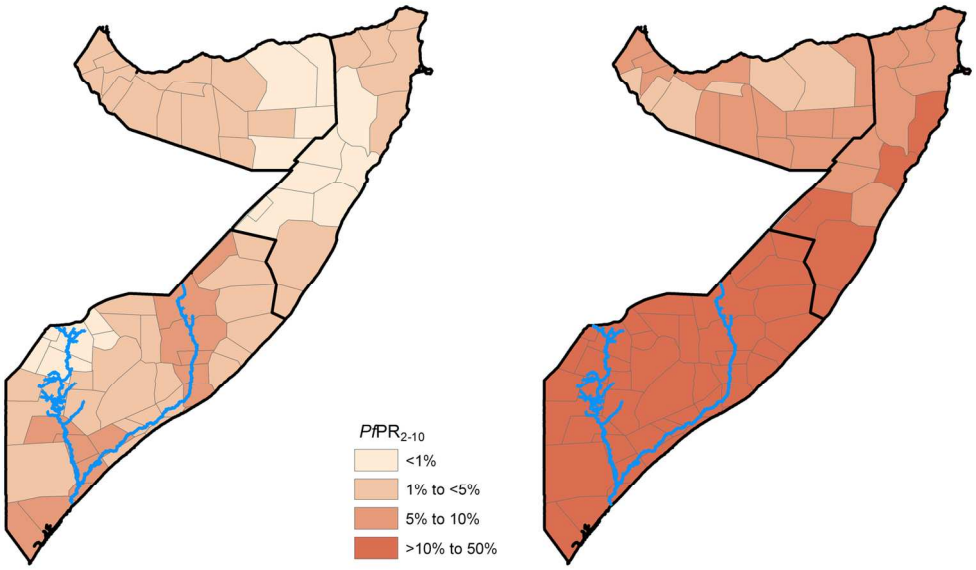
	Endemicity classification based on the 2010 annual mean *Pf*PR_{2-10} (contemporary) predictions		**Endemicity classification based on the maximum mean *Pf*PR_{2-10} (receptive) predictions over the period 2007-2010**	
	Number (%) of districts	**Population at risk, million, (%)**	**Number (%) of districts**	**Population at risk, million, (%)**
Population weighted mean *Pf*PR_{2-10}				
Hypoendemic				
<1% *Pf*PR_{2-10}	17 (23)	1.1 (13)	0 (0)	0 (0)
1% to <5% *Pf*PR_{2-10}	43 (58)	4.6 (55)	6 (8)	1.2 (14)
5% to 10% *Pf*PR_{2-10}	14 (19)	2.6 (31)	20 (27)	1.4 (17)
Mesoendemic (>10% to 50% *Pf*PR_{2-10})	0 (0)	0.0 (0)	48 (65)	5.8 (69)
Hyperendemic and Holoendemic (>50% *Pf*PR_{2-10})	0 (0)	0.0 (0)	0 (0)	0.0 (0)



Zone, regional and district maps of Somalia showing the distribution of the age-standardised community Plasmodium falciparum parasite rate (PfPR2-10) data (n=1,558) assembled during the period 2007 – 2010 (including 54 surveys undertaken in 2011). The zones are CS= Central South; NE= North East; NW= North West. The thick black line show the zone boundaries, the thin black lines show the regional boundaries and the thin grey lines show the district boundaries. The blues lines show the location of the Juba (lower) and Shabelle (upper) Rivers.
215x279mm (96 x 96 DPI)



A) Map of the posterior annual mean PfPR2-10 prediction to 2010 (contemporary) at 1×1 km grid location in Somalia; B) Map of the maximum mean PfPR2-10 prediction (receptive) at 1×1 km grid location as computed from the posterior annual mean PfPR2-10 prediction for each year from 2007 to 2010; C) Map of the coefficient of variation (the standard deviation/the mean PfPR2-10 prediction) of the contemporary prediction at 1x1 km grid location; D) Map of the coefficient of variation at 1x1 km grid location of the receptive prediction. The thick black lines show the zone boundaries, the thin black lines show the regional boundaries and the thin grey lines show the district boundaries. Higher coefficient of variation of the predictions suggests higher uncertainties of the PfPR2-10 predictions. The scale bar for the continuous PfPR2-10 ends at 50 which is the upper limit of mesoendemic transmission. The blues lines show the location of the Juba (lower) and Shabelle (upper) Rivers.
253x321mm (300 x 300 DPI)



District (n=74) maps of Somalia classified by endemicity using a population-weighted: A) posterior aggregate annual mean PfPR2-10 (contemporary) prediction to 2010 and; B) the maximum annual mean PfPR2-10 (receptive) predictions over the period 2007-2010. The blues lines show the location of the Juba (lower) and Shabelle (upper) Rivers.
148x89mm (300 x 300 DPI)

Supplementary Information (SI) 1: Predictors of age-standardised *Plasmodium falciparum* parasite rate (PfPR$_{2-10}$)

SI 1.1 Urbanisation: A surface of urbanisation in Somalia derived from a 100×100 m spatial resolution population surface developed from a combination of census, satellite imagery and land cover data[1,2] was used to define each survey location as urban or rural. Somalia population map was derived by a combination of simple and semi-automated processes involving the re-distribution of district level population estimates to a finer spatial scale using land cover and land use datasets[1]. A refined land cover map was prepared by combining the settlement data with Africover data[3] and other ancillary GIS data on roads, rivers from Vector Map Level Zero (VMAP 0)[4] and the Landsat satellite imagery. A total of 22 land cover classes were formed and population density within these classes was calculated with urban and rural extents derived from the GEOterrain counsultancy[5]. These calculated densities were then used as weightings to redistribute population by settlement and in land cover types that were unaccounted for by existing settlement size data (Figure SI 1A).

SI 1.2 Enhanced vegetation Index (EVI): EVI is an index of intensity of photosynthetic activity and a good proxy for rainfall ranges from 0 (no vegetation) to 1 (complete vegetation). Monthly EVI surfaces at 1×1 km spatial resolution derived from the global Moderate Resolution Imaging Spectro-radiometer (MODIS) satellite imagery for the period 2001-2010 were downloaded from the MODIS webiste[6] and were used to compute annual mean EVI. These monthly maps were used to compute annual mean EVI for the years 2007 to 2010 (Figure SI 1B).

SI1.3 Precipitation: Monthly mean precipitation (mm) raster surfaces at 1×1 km resolution were downloaded from the WorldClim website[7] and used as a proxy for rainfall. The Worldclim database was compiled using weather stations data collected world-wide for the period 1950-2000 and interpolated using spline methods[8]. The monthly mean precipitation was used to compute annual mean precipitation surfaces for the years 2007 to 2010 (Figure SI 1C).

SI1.4 Temperature suitability Index (TSI): As metric for the effect of temperature on malaria transmission, TSI at a spatial resolution of 1×1 km[9] was used. TSI was constructed using monthly temperature time series[7] within a biological modelling framework to quantify the effect of ambient temperature on sporogony and vector survivorship and determine the suitability of an area to support transmission globally[9]. On a scale of increasing transmission suitability, TSI ranges from 0 (unsuitable) to 1 (most suitable) (Figure SI 1.1.1D).

SI 1.5 Potential mosquito breeding sites: To define potential breeding sites the Global Lakes and Wetlands databases (GLWD) developed as a partnership between the World Wildlife Fund and the Center for Environmental Systems Research, University of Kassel, Germany[10] was used. The GLWD data contained 12 wetland classifications but for Somalia the following wetland classes were retained for analysis: lake, rivers, floodplains, reservoirs and coastal wetlands. Intermittent water bodies and brackish and saline water were removed. Euclidean distances to these wetlands were computed in ArcGIS 10 (ESRI Inc. NY, USA) (Figure SI 1E).

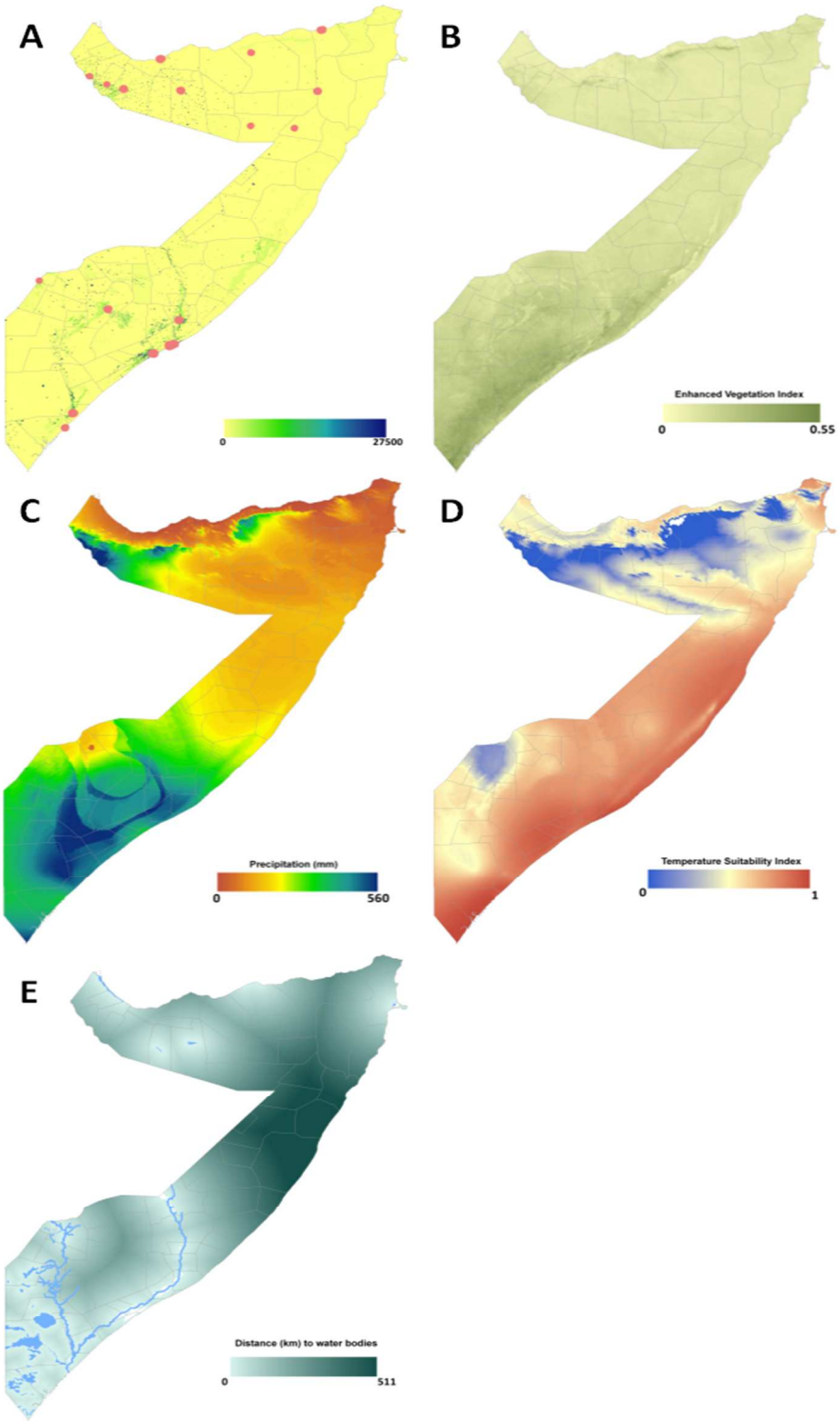
SI 1.6 Statistical analysis of the predictors of $PfPR_{2-10}$

The values of the underlying ecological, climatic and population covariates describe above were extracted to each survey location using ArcGIS 10 *Spatial Analyst* tool. Distance to potential breeding sites was log-transformed before analysis because of its high positive skew. The covariates were then included in total-sets analysis which is an automatic model selection process based on a generalized linear regression model and implemented using the *bestglm* package in R[11,12]. This approach selects the best combination of the covariates based on the value of the Bayesian Information Criteria (BIC) statistic[13] which selects the lowest BIC as the best model fit. The model excluding TSI showed the lowest BIC and therefore the best fit (Table SI 1.1.1).

Table SI 1.1.1 Summary of regression analysis of covariates

Covariate	Odds Ratio (95% CI)	P-value
Urbanisation	-0.88 (-1.02- -0.69)	<0.001
EVI	0.81 (0.19-1.44)	0.011
Precipitation	0.003 (0.002-0.004)	<0.001
TSI	-0.68 (-0.89- 0.01)	<0.312
Log of distance to wetlands	0.27 (0.24 – 0.31)	<0.001

Figure SI 1 Somalia maps of 1×1 km spatial resolution of: **A)** population distribution showing the location of urban centres (in red); **B)** annual mean enhanced vegetation index (EVI); **C)** annual mean precipitation (mm); **D)** temperature suitability index for *Plasmodium falciparum* transmission; **E)** distance (km) to nearest water body



References

1. Linard C, Alegana VA, Noor AM, Snow RW *et al.,* A high resolution spatial population database of Somalia for disease burden estimation. *International Journal of Health Geographics* 2010; 9: 45.
2. The AfriPop Project. http://www.clas.ufl.edu/users/atatem/index_files/Details.htm. Accessed 15 January 2012.
3. Africover. www.africover.org. Accessed 13 February 2012.
4. VMAP 0. http://www.mapability.com/index1.html?http&&&www.mapability.com/info/vmap0_index.html. Accessed 3 January 2012.
5. GeoTerraImage. www.geoterraimage.com. Accessed 15 February 2012.
6. Nasa TRMM. Goddard, Maryland, USA, NASA Goddard Space Flight Center. http://trmm.gsfc.nasa.gov/. Accessed 15 November 2011.
7. WorldClim. http://www.worldclim.org/download.htm. Accessed 10 December 2011.
8. Hijmans RJ, Cameron SE, Parra JL, Jones PG *et al.*.Very high resolution interpolated climate surfaces for global land areas. *Int J Climatology* 2005; 25: 1965-1978.
9. Gething PW, Van Boeckel T, Smith DL, *et al.* Modelling the global constraints of temperature on transmission of *Plasmodium falciparum* and *P. vivax*. *Parasite Vector* 2011; 4:92.
10. Global Lakes and Wetlands Database Request. https://secure.worldwildlife.org/science/data/item1877.html. Accessed 10 September 2011.
11. Miller A. *Subset Selection in Regression.* Boca Raton, FL: Chapman & Hall. 2002:238 p.
12. Lumley T. *leaps: regression subset selection* (R package) Version 2.7, 2010.
13. Schwarz G. Estimating dimensions of a model. *Ann Stat* 1978; 6: 461-464.

Supplementary Information (SI) 2: Model-based Geostatistical Procedures

Below are the details of the MBG framework used to develop the contemporary and receptive malaria risk maps for Somalia. The MBG procedure was adopted from Gething et al 2011[1] where further model details, especially on age-standardisation of parasite rate data, can also be found. The generic model code is available on Malaria Atlas Project *P. falciparum* Cartographic code link[2] and has been adapted for Somalia. Model fitting was achieved using Markov chain Monte Carlo (MCMC)[3,4].

SI 2.1 The MBG presentation

Each of the N_i individuals in sample i was assumed *P. falciparum* positive with probability $\tilde{k}_i P'(x_i, t_i)$, so the number positive N_i^+ was distributed binomially:

$$N_i^+ | N_i, P'(x_i, t_i) \overset{\text{ind}}{\sim} \text{Bin}(N_i, \tilde{k}_i P'(x_i, t_i)) \qquad \text{S2.1}$$

The coefficient $P'(x_i, t_i)$ was modelled as a Gaussian process. The factor $\tilde{k}_i$ converted $P'(x_i, t_i)$ to the probability that individuals within the age range reported for study i were *P. falciparum* positive, and that the infection was detected, thereby accounting for the influence of age on the probability of detection[5]. The age-standardisation factor $\tilde{k}_i$ in each population was assumed drawn independently from a distribution $D_{\tilde{k}}$ whose parameters were the lower $A_{L,i}$ and upper $A_{U,i}$ ages reported in study i:

$$\tilde{k}_i | A_{U,i}, A_{L,i} \overset{\text{ind}}{\sim} D_{\tilde{k}}(A_{U,i}, A_{L,i}) \qquad \text{S2.2}$$

The form of $D_{\tilde{k}}$ is described in Gething et al 2011[2].

*Pf*PR$_{2\text{-}10}$ is the *P. falciparum* parasite rate for individuals between ages 2 (2.00) and 10 (9.99). Its value at an arbitrary location x and time t is the product of $P'(x,t)$ and another age-standardisation factor, k_{2-10}, distributed as $D_k(2, 10)$:

$$\begin{aligned} &\text{PR}_{2-10}(x,t) = P'(x,t) k_{2-10}(x,t) \\ &k_{2-10}(x,t) \overset{\text{ind}}{\sim} D_k(2,10) \end{aligned} \qquad \text{S2.3}$$

The factor k_{2-10} converted $P'(x,t)$ to the probability that individuals between ages 2 and 10 at location x are *P. falciparum* positive. The age-standardisation factor $\tilde{k}$ of a survey is the product of the age-standardisation factor k associated with the same place, time and age range and the sensitivity of the survey.

The coefficient $P'(x,t)$ at arbitrary location x and time t was modelled as the inverse-logit function applied to a random field f evaluated at (x,t), plus an unstructured (random) component $\epsilon(x,t)$.

$$P'(x,t) = \text{logit}^{-1}(f(x,t) + \epsilon(x,t)) \quad \text{S2.4}$$

The components $\epsilon(x,t)$ were assumed independent and identically distributed for each location x and time t and a standard diffuse but proper prior with expectation 0.25 was assigned to their variance V.

$$\epsilon(x,t)|V \overset{\text{iid}}{\sim} \text{N}(0,V) \quad \text{S2.5}$$

$$\frac{1}{v} \sim Gamma(3,12) \quad \text{S2.6}$$

The random field f was modelled as a Gaussian process characterised by its mean and covariance functions:

$$f(x,t)|\boldsymbol{\beta},\tau,\phi_x,\phi_t,\lambda,\psi,\rho,\upsilon \sim \text{GP}(\boldsymbol{\beta},C) \quad \text{S2.7}$$

The mean function was defined as $\mu = \boldsymbol{\beta}\mathbf{X}$,where $\mathbf{X} = 1, X_1(x), ..., X_n(x)$ was a vector consisting of a constant and $n = 20$ environmental covariates indexed by spatial location x ,and $\boldsymbol{\beta} = \beta_0, \beta_1, ..., \beta_n$ was a corresponding vector of regression coefficients. The covariance of the field was modelled using a version of the spatiotemporal covariance function recently recommended by Stein[6] (equation 2.12):

$$C(x_i,t_i;x_j,t_j) = \tau^2\gamma(0)\frac{(\Delta x)^{\gamma(\Delta t)}K_{\gamma(\Delta t)}(\Delta x)}{2^{\gamma(\Delta t)-1}\Gamma(\gamma(\Delta t)+1)},$$

$$\gamma(\Delta t) = \frac{1}{2\rho+2(1-\rho)\left[(1-\upsilon)e^{-|\Delta t|/\phi_t}+\upsilon\cos(2\pi\Delta t)\right]}, \quad \text{S2.8}$$

$$\Delta t = |t_i - t_j|$$

K_γ is the modified Bessel function of the second kind of order γ, and Γ is the gamma function [7,8].

Spatial distance between a pair of points x_i and x_j was computed as great-circle distance $D_{GC}(x_i,x_j)$ multiplied by a factor that depends on the angle of inclination $\theta(x_i,x_j)$ of the vector pointing from x_i to x_j. θ was computed as if latitude and longitude were Euclidean coordinates (on a cylindrical projection) to allow for anisotropy:

$$\Delta x = 2\sqrt{\gamma(\Delta t)}\frac{D_{GC}(x_i, x_j)\sqrt{1 - \psi^2 \cos^2(\theta(x_i, x_j) - \lambda)}}{\phi_x} \quad \text{S2.9}$$

When $\Delta x = 0$ (that is, for points at the same location but different times), the covariance function reduces to

$$\rho + (1 - \rho)\left[(1 - \upsilon)e^{-|\Delta t|/\phi_t} + \upsilon \cos(2\pi\Delta t)\right] \quad \text{S2.10}$$

As temporal separation increases, the covariance approaches a limiting sinusoid $\tau^2[\rho + (1 - \rho)\upsilon \cos(2\pi\Delta t)]$ rather than zero. When $\Delta t = 0$, on the other hand (for points at different locations but the same time), it reduces to a standard exponential form with range parameter $\phi_x\sqrt{2}$. Unlike standard sum-product models, this covariance function does not have problematic ridges along its axes[6].

SI 2.2 Prior Specification

The square root of the partial sill τ and the spatial range parameter ϕ_x were assigned skew-normal priors:

$$\log \tau | \mu_\tau, V_\tau, \alpha_\tau \sim \text{Skew-Normal}(\mu_\tau, V_\tau, \alpha_\tau) \quad \text{S2.11}$$

$$\log \phi_x | \mu_\phi, V_\phi, \alpha_\phi \sim \text{Skew-Normal}(\mu_\phi, V_\phi, \alpha_\phi) \quad \text{S2.12}$$

and their specification is described further below.

The standard "one-over-*x*" prior for the temporal scale parameter ϕ_t resulted in collapse to zero, a common artefact when data do not contain strong information. A relatively vague but proper prior, which has an expectation of ten years, was used instead.

$$\phi_t \sim \text{Exponential}(0, .1) \quad \text{S2.13}$$

A uniform prior was assigned to the direction of anisotropy parameter λ and to the square of the "eccentricity" parameter ψ, which controls the amount of anisotropy,

$$\lambda \sim \text{Uniform}(0, \pi) \quad \text{S2.14}$$

$$\psi^2 \sim \text{Uniform}(0, 1) \quad \text{S2.15}$$

a uniform prior was assigned to the temporal parameters governing the amplitude of the sinusoidal component ρ and the limiting autocorrelation in the temporal direction υ:

$$\rho \sim \text{Uniform}(0,1), \upsilon \sim \text{Uniform}(0,1) \qquad \text{S2.16}$$

and a standard prior was assigned to the components of the mean:

$$p(\boldsymbol{\beta}) \propto 1 \qquad \text{S2.17}$$

Although standard priors such as the improper "flat" prior[3] were assigned to most of the basic model parameters, subjective skew-normal priors[7] were specified for the range and partial sill parameters τ and ϕ_x.

SI 2.3 Model implementation

SI 2.3.1 MCMC Algorithms

Both the main geostatistical model and the age-standardisation sub-model were fitted using the MCMC algorithm[3,4]. The algorithm was implemented in the Python[8] and FORTRAN programming languages using the open-source Bayesian statistics package PyMC[9,10] and the numerical packages SciPy and NumPy[11].

The evaluation of f at the sampling locations and times was updated using Gibbs steps[3]. The evaluation of the uncorrelated process ϵ was updated one point at a time using random-walk Metropolis steps[3]. The model parameters $\boldsymbol{\beta}$, τ, ϕ_x, ϕ_t, λ, ψ, V and ρ were updated jointly using the method of Haario, Saksman and Tamminen[12].

Within the MCMC loop, the age-standardisation factors $\tilde{k}_i$ were not imputed explicitly. We were not interested in their particular values, and marginalizing out "nuisance parameters" ahead of time usually improves the mixing of MCMC algorithms. Before the MCMC loop began, the marginal likelihood:

$$\int \text{Bin}(N_i^+; N_i, k_i P'(x_i,t_i)) D_{\tilde{k}}(\tilde{k}_i; A_{U,i}, A_{L,i}) d\tilde{k}_i \qquad \text{S2.18}$$

was approximated using standard Monte Carlo integration for several values of $P'(x_i,t_i)$. That is, values for the model parameters α_i, b_i, c_i and s_i and the age distribution S_i were drawn from their posterior predictive distributions, then expression (S3.1) was evaluated to obtain k_i, then the binomial probability

was evaluated for several values of $P'(x_i, t_i)$. The probabilities resulting from many such draws were averaged. Inside the MCMC loop, the marginal likelihood function for arbitrary values of $P'(x_i, t_i)$ was evaluated by interpolation.

SI 2.3.2 Age Correction Model

The age distribution parameters S_i, S_0 and ν are independent of the relative *Pf*PR parameters P_i', α_i, c_i, b_i, s_i, μ_A, σ and R given the data, so these two groups of parameters were inferred using separate MCMC algorithms.

In the MCMC for the age distribution parameters, the survey populations' age distributions S_i were updated using Gibbs steps[3]. The concentration parameter ν was updated using random-walk Metropolis steps[3]. The typical age distribution S_0 was represented as a normalized sequence of gamma random variables[13], and these variables were updated one at a time using random-walk Metropolis steps[3].

In the MCMC for the relative *Pf*PR parameters, the distributional parameters μ_A, σ and R were updated jointly using the method of Haario, Saksman and Tamminen[12]. The parameters P_i', α_i, c_i, b_i and s_i were updated jointly for each population i using the same method.

SI 2.3.3 Spatiotemporal Prediction

The output of the MCMC stage consisted of $\{\theta_{(l)}; l = 1, ..., m\}$ samples from the posterior of the parameter set $\theta = \{\boldsymbol{\beta}, \tau, \phi_x, \phi_t, \lambda, \psi, \rho, k, V\}$ and a corresponding $\{f(x_i, t_i)_{(l)}; l = 1, ..., m\}$ samples from the posterior of the space-time random field at each of the n data locations$\{(x_i, t_i); i = 1, ..., n\}$. For every $l'th$ sample, the conditional distribution of the annual mean of the space-time random field was predicted at each prediction location x_j on the nodes of a regular 1×1 km grid within the spatial limits of stable *P. falciparum* transmission[14]. The distribution of the annual mean $f(x_j)_{(l)}$ for prediction location x_j was modelled as the joint multivariate normal distribution of the 12 predicted monthly values $\{t = 2007_{Jan}, ..., 2007_{Dec}\}$ for that year specified by a 12 element mean vector $\hat{\boldsymbol{y}}(x_j)_{(l)}$ and 12 × 12 variance-covariance matrix $\hat{\boldsymbol{\sigma}}^2(x_j)_{(l)}$:

$$f(x_j)_{(l)} \sim MVN\left(\hat{\boldsymbol{y}}(x_j)_{(l)}, \hat{\boldsymbol{\sigma}}^2(x_j)_{(l)}\right) \quad \text{S2.19}$$

The mean vector $\hat{\boldsymbol{y}}(x_j)_{(l)}$ was computed using:

$$\hat{\boldsymbol{y}}(x_j)_{(l)} = \boldsymbol{\mu}_{P_{(l)}} + \boldsymbol{C}^T_{DP_{(l)}} \cdot \boldsymbol{C}^{-1}_{DD_{(l)}} \cdot (\boldsymbol{p}(x,t) - \boldsymbol{\mu}_{D_{(l)}}) \quad \text{S2.20}$$

where $\boldsymbol{\mu}_P$ and $\boldsymbol{\mu}_D$ were the predicted mean of the random field at each of the 12 prediction times $\{t = 2007_{Jan}, ..., 2007_{Dec}\}$ at spatial location x_j and at each of the n data locations respectively, $\boldsymbol{C}_{DP}$ and $\boldsymbol{C}_{DD}$ were the data-to-prediction and data-to-data covariance matrices respectively, and $\boldsymbol{p}(x,t)$ was the vector of n data values. The 12 × 12 variance-covariance matrix $\hat{\boldsymbol{\sigma}}^2(x_j)_{(l)}$ was computed using:

$$\hat{\boldsymbol{\sigma}}^2(x_j)_{(l)} = \boldsymbol{C}_{PP_{(l)}} - \boldsymbol{C}^T_{DP_{(l)}} \cdot \boldsymbol{C}^{-1}_{DD_{(l)}} \cdot \boldsymbol{C}_{DP_{(l)}} \quad \text{S.2.21}$$

The value of the $l'th$ sample of V, the variance of the unstructured component $\epsilon(x,t)$, was then added to the diagonal of the matrix $\hat{\boldsymbol{\sigma}}^2(x_j)_{(l)}$ and 1000 draws were made randomly from the distribution specified in equation S2.34. These draws represented samples from the posterior distribution of $f(x_j)$ and were subject to an inverse logit transform and then multiplied by the $l'th$ sample of the age-standardisation parameter $k_{2-10(l)}$ to form the $l'th$ sample from the posterior distribution of the predicted mean annual 2010 *Pf*PR$_{2\text{-}10}$ endemicity surface at location x_j:

$$P'_{2-10}(x_j)_{(l)} = \text{logit}^{-1}\left(f(x_j)_{(l)} + \epsilon_{(l)}\right) k_{2-10(l)} \quad \text{S.2.22}$$

This procedure was repeated for every $l'th$ sample to form the set $\{P'_{2-10}(x_j)_{(l)}; l = 1, ..., m\}$ of m samples for each prediction location. The point estimate of *Pf*PR$_{2\text{-}10}$ endemicity at each location was defined as the mean of this set, whilst the probability of membership to each class was computed as the proportion of these samples falling within each class definition.

References

1. Gething PW, Patil AP, Smith DL, *et al*. A new world malaria map: Plasmodium falciparum endemicity in 2010. *Malar J* 2011; 10: 378 doi: 10.1186/1475-2875-10-378.
2. Malaria Atlas Project *P. falciparum* Cartographic code. https://github.com/malaria-atlas-project/mbgw-clean. Accessed December 2011.
3. Gilks WR, Spiegelhalter DJ. Markov Chain Monte Carlo in practice. Interdisciplinary Statistics. Boca Raton, Florida, U.S.A.: Chapman & Hall / CRC Press LLC, 1999.
4. Gelman A, Carlin JB, Stern HS. Bayesian data analysis. Texts in Statistical Science. Boca Raton, Florida, U.S.A.: Chapman & Hall / CRC Press LLC. 696 p, 1993.
5. Smith DL, Guerra CA, Snow RW, Hay SI. Standardizing estimates of the *Plasmodium falciparum* parasite rate. *Malar J* 2007; 6: 131.

6. Stein ML. Space-time covariance functions. *J Am Stat Assoc* 2005; 100: 310-321.

7. Azzalini A. A class of distributions which includes the normal ones. *Scand J Stat* 1985; 12: 171-178.

8. van Rossum G (2008) Python Programming Language - Official Website. Website: URL http://www.python.org.

9. Fonnesbeck C, Huard D, Patil AP. PyMC 2.0 User's Guide: installation and tutorial 2008. URL: http://www.trichech.us/pymc.

10. Patil A, Huard D, Fonnesbeck CJ. PyMC: Bayesian stochastic modelling in Python. *J Stat Softw* 2010; 35: e1000301.

11. Oliphant TE. Python for scientific computing. *Comput Sci Eng* 2007; 9: 10-20.

12. Haario H, Saksman E, Tamminen J. An adaptive Metropolis algorithm. *Bernoulli* 2001; 7: 223-242.

13. Hogg RV, Craig AT. *Introduction to Mathematical Statistics*. Upper Saddle River, New Jersey, U.S.A: Prentice Hall Inc. 2005; 564 p.

14. Guerra CA, Gikandi PW, Tatem AJ, Noor AM, Smith DL, *et al.* The limits and intensity of *Plasmodium falciparum* transmission: implications for malaria control and elimination worldwide. *PLoS Med* 2008; 5: e38.

Mapping the receptivity of malaria risk to plan the future of control in Somalia

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2012-001160.R1
Article Type:	Research
Date Submitted by the Author:	08-Jun-2012
Complete List of Authors:	Noor, Abdisalan; Kenyan Medical Researh Institute/University of Oxford-Wellcome Trust Research Programme, Malaria Public Health & Epidemiology Group Alegana, Victor; Kenyan Medical Researh Institute/Wellcome Trust Research Prgramme, Malaria Public Health & Epidemiology Group, Nairobi, Kenya Patil, Anand; Sense Inc., Moloney, Grainne; United Nations Food and Agricultural Organization, Food Security and Nutrition Analysis Unit-Somalia Borle, Mohammed; United Nations Food and Agricultural Organization, Food Security and Nutrition Analysis Unit-Somalia Yusuf, Fahmi; World Health Organization, Roll Back Malaria, Somalia Office Amran, Jamal; World Health Organization, Roll Back Malaria, Somalia Office Snow, Robert; Kenyan Medical Researh Institute/University of Oxford-Wellcome Trust Research Programme, Malaria Public Health & Epidemiology Group
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Public health, Infectious diseases, Health policy
Keywords:	Epidemiology < INFECTIOUS DISEASES, Public health < INFECTIOUS DISEASES, Geographical mapping < TROPICAL MEDICINE

Mapping the receptivity of malaria risk to plan the future of control in Somalia

Abdisalan M Noor [1,2*], Victor A Alegana[1], Anand P Patil[3], Grainne Moloney[4], Mohammed Borle[4], Fahmi Yusuf[5], Jamal Amran[5], Robert W Snow[1,2]

1. Malaria Public Health & Epidemiology Group, Kenya Medical Research Institute/Wellcome Trust Research Programme, Centre for Geographic Medicine Research-Coast, P.O. Box 43640, 00100 GPO, Nairobi, Kenya.

2. Nuffield Department of Medicine, Centre for Tropical Medicine, University of Oxford, John Radcliffe Hospital, Headington, Oxford, OX3 9DU, UK

3. Sense Inc., 90 Lawrence St., Detroit, MI, USA, 48202.

4. United Nations Food and Agricultural Organization, Food Security and Nutrition Analysis Unit-Somalia, 3rd Floor P.O. Box 1230, Village Market, Nairobi, Kenya

5. World Health Organization, Malaria Control & Elimination, Somalia.

***AMN:** anoor@nairobi.kemri-wellcome.org
VAA : valegana@nairobi.kemri-wellcome.org
APP : anand.prabhakar.patil@gmail.com
GM : Grainne.Moloney@fao.org
MB: Mohammed.Borle@fao.org
FY: yusuff@som.emro.who.int
JA : amranj@nbo.emro.who.int
RWS: rsnow@nairobi.kemri-wellcome.org

*Author for correspondence

ABSTRACT

Objectives: To measure the receptive risks of malaria in Somalia and compare decisions on intervention scale-up based on this map and the more widely used contemporary risk maps.

Design: Cross-sectional community *Plasmodium falciparum* parasite rate (*Pf*PR) data for the period 2007 to 2010 corrected to a standard age-range of two to less than 10 years ($PfPR_{2-10}$) and used within a Bayesian space-time geostatistical framework to predict the contemporary (2010) mean $PfPR_{2-10}$ and the maximum annual mean $PfPR_{2-10}$ (receptive) from the highest predicted $PfPR_{2-10}$ value over the study period as an estimate of receptivity.

Setting: Randomly sampled communities in Somalia.

Participants: Randomly sampled individuals of all ages.

Main outcome measure: Cartographic descriptions of malaria receptivity and contemporary risks in Somalia at the district level.

Results: The contemporary annual $PfPR_{2-10}$ map estimated that all districts (n=74) and population (n=8.4 million) in Somalia were under hypoendemic transmission (≤ 10% $PfPR_{2-10}$). Of these, 23% of the districts, home to 13% of the population, were under transmission of <1% $PfPR_{2-10}$. About 58% of the districts and 55% of the population were in the risk class of 1% to <5% $PfPR_{2-10}$. In contrast the receptivity map estimated 65% of the districts and 69% of the population were under mesoendemic transmission (>10% to 50% $PfPR_{2-10}$) and the rest as hypoendemic.

Conclusion: Compared to maps of receptive risks, contemporary maps of transmission mask disparities of malaria risk necessary to prioritize and sustain future control. As malaria risk declines across Africa, efforts must be invested in measuring receptivity for efficient control planning.

Article focus

- Cross-sectional *PfPR* prevalence survey data for the period 2007 to 2010 in Somalia
- Bayesian geostatistical models estimating the receptive and contemporary malaria transmission in Somalia
- Implications of the two malaria risk maps for malaria control planning in Somalia

Key messages

- It is feasible to use *Pf*PR community prevalence data in space and time to estimate the receptive and contemporary risks of malaria within the same model framework.
- Malaria receptivity maps are critical to inform the scale-up and sustaining of interventions where disease has declined or is highly seasonal.
- Efforts must be invested in helping malaria endemic countries in Africa measure their receptive risks.

Strengths and limitations

- The annual *Pf*PR surveys provide unique opportunities to measure receptivity in Somalia.
- Improving the spatial and temporal distributions of *Pf*PR data and exploring probabilistic approaches of selecting potential maximum risks will improve the measurement of receptivity.

INTRODUCTION

Malaria receptivity is a measure of the intrinsic vector transmission potential of an area[1]. Interest in measuring malaria receptivity has emerged following the resurgence of the malaria elimination agenda[2,3] and the need to quantify the risks posed by human population movement leading to the reintroduction of transmission[4,5]. However, understanding receptivity is equally important to decision-making for countries that are implementing control. In low stable endemic countries national programmes need to understand the risks posed by withdrawal of interventions from areas that are historically high transmission[3,6]. In unstable transmission areas where parasite exposure is highly seasonal and prone to climatic anomalies, targeting interventions to prevent the risk of epidemics are a priority[1]. The empirical malaria risk maps that are commonly available to countries to support malaria control planning are those that represent the contemporary distribution of risk under control[7,8] and are therefore of limited value in defining the epidemic potential or the receptive rebound risks of withdrawing or failing to sustain interventions.

Measuring receptivity for malaria control countries ideally requires empirical data from a period of no control and under the optimum transmission conditions. There is hardly any country in Africa that has remained universally control-naive over the last 100 years[9]. Alternatively, nationally representative empirical data on malaria transmission during the pre-Roll Back Malaria (RBM) era, or before intervention scale-up reached critical thresholds for a given country, may represent the best approximation of receptivity. Furthermore, such information must be resolved to administrative decision-making units to which malaria resources are allocated to make them relevant for policy. In this study we use community *Plasmodium falciparum* parasite prevalence data from 2007 to 2010 within a model-based geostatistical (MBG) framework to develop contemporary and receptive risk maps and resolve endemicity to districts in Somalia.

METHODS

Country context

Somalia is divided into the three zones of North West (Somaliland), North East (Puntland) and Central South (Figure 1). The northern zones are generally dry and hot whereas the Central South zone has sub-tropical climate and is where the two major rivers of the country, the Shabelle and the Juba, are located[10-11]. *Anopheles arabiensis* is the dominant malaria transmitting vector throughout the country although *An. funestus* is reported in Central South[12, 13]. *P. falciparum* is the dominant species of the malaria parasite[14-16]. The presence of *P. vivax* cases have also been reported with studies in Somaliland showing relatively high vivax antibody responses[17]. The failure of the long rains in Somalia in 2010 combined with the below average rainfalls in previous two seasons have resulted in a severe drought[11].

Since 1991 Somalia has had no effective central government and has experienced frequent internal armed conflicts resulting in the breakdown of public services and political disintegration[18]. Although the transitional federal government (TFG), formed in February 2004, is internationally recognized as the official government of Somalia[19], in reality the three zones currently function as semi-independent states. Consequently, there are three ministries of health but the majority of health care is provided by international and national non-government organizations that have come together under the umbrella of Somalia Aid Coordinating Board which was later reformed as the Somalia Support Secretariat[20].

The main funder of malaria control in Somalia is the Global Fund to Fight Aids TB and Malaria (GFATM) with the United Nations Children Fund (UNICEF) as the principal recipient and the World Health Organization (WHO) as sub-recipient[21,22]. In 2010, the second national malaria strategy was launched with universal scale-up of vector control, parasitological diagnosis, effective antimalarials, improved surveillance and epidemic preparedness and response as the main strategic approaches[23]. Since 2004, over 77 million USD has been approved to Somalia by the GFATM for malaria control resulting in the distribution of almost one million LLINs[24].

Community survey data

The community *P. falciparum* parasite rate (*Pf*PR) is the most commonly used indicator for mapping malaria transmission[7]. This is because it is easy to measure, has a historical legacy and a predictable relationship with other measures of transmission intensity such as entomological inoculation rate and the basic reproductive rate[25,26]. The *Pf*PR data used for the present study were assembled through the Food and Agriculture Organization-Food Security and Nutrition Analysis Unit (FAO-FSNAU) surveys undertaken regularly in

Somalia[16,27]. These surveys were initially established to monitor the nutritional status of children less than 5 years of age and began in 2000[27]. Investigations of malaria prevalence covering persons of all ages were only included from 2007[16] and have been undertaken annually since, covering most regions of Somalia. A detailed description of the sampling design is provided elsewhere[16]. During the survey, respondents provided a finger prick blood sample that was examined for the presence of *P. falciparum* infection using a rapid diagnostic test (RDT) (Paracheck Pf TM, Orchid Biomedical Systems, Goa, India). Consent was obtained for all individuals before interview and separately for the malaria testing. Additional information was recorded on the date of survey and age and sex of participants. After all survey data were assembled each surveyed community was geo-coded using combinations of global positioning systems (GPS), electronic gazetteers (Google Earth, Encarta and Alexandria), and other sources of longitude and latitude such a settlement database collated by FAO-SWALIM[11].

Assessment of ecological and climatic predictors of malaria risk

The transmission intensity of malaria is influenced by climatic and ecological factors through their independent or combined effects on the survival of the Anopheline vectors and the Plasmodium parasites within the vector[28]. The ecological and climatic factors used commonly to improve the precision of empirical malaria mapping are urbanisation, rainfall, temperature and distance to potential mosquito larva breeding sites. ~~or their proxies.~~

Data at 1 × 1 km spatial resolution on urbanisation[19,30], annual mean precipitation[31,32] and enhanced vegetation index (EVI)[33] were assembled. Maps of rivers, floodplains, reservoirs and coastal wetlands were assembled from the Global Wetlands and Lakes Database[34] and Euclidean distances to these proximates of breeding sites were computed in ArcGIS 10 (ESRI Inc. NY, USA). As a metric for the effect of temperature on malaria transmission, a temperature suitability index (TSI) at a spatial resolution of 1 × 1 km was used. TSI was constructed using monthly temperature time series within a biological modelling framework to quantify the effect of ambient temperature on sporogony and vector survivorship and determine the suitability of an area to support transmission globally[35]. The values of the underlying ecological and climatic covariates were extracted to each survey location using ArcGIS 10 *Spatial Analyst* tool. Distance to potential breeding sites was log-transformed before analysis because of its high positive skew. The covariates were then included in total-sets analysis which is an automatic model selection process based on a generalized linear regression model and implemented using the *bestglm* package in R[36,37]. This approach selects the best combination of the covariates based on the value of the Bayesian Information Criteria (BIC) statistic[38] which selects the lowest BIC

as the best model fit. Details of the ecological and climatic predictors of *Pf*PR and the results of the total-set analysis are provided in Supplementary Information (SI) 1.

The space-time Bayesian geostatistical model for predicting *P. falciparum* distribution in Somalia

Space-time MBG methods offer the flexibility of predicting an outcome to any given year in a time series by harnessing fully both the spatial and temporal relationships of the data and generate uncertainties of the predictions from the full posterior distributions[39].

In this study, the assembled *Pf*PR data were standardized to the classical age-range of 2 to less than 10 years using an algorithm based on modified catalytic conversion models[40]. The continuous surfaces of the age-standardised data ($Pf\text{PR}_{2\text{-}10}$) were generated using a space-time MBG framework[8,41] whereby Bayesian inference was implemented using the Markov Chain Monte Carlo algorithm. Details of model code[41] and statistical procedures[8] are provided in SI 2. In brief, the value of $Pf\text{PR}_{2\text{-}10}$ was modelled as a transformation of a spatiotemporally structured field superimposed with unstructured (random) variation on a regular 1 × 1 km grid from 2007 to 2010. The number of *P. falciparum* positive responses from the total sample at each survey location was modelled as a conditionally independent binomial variate given the unobserved underlying age-standardised $Pf\text{PR}_{2\text{-}10}$ value[40] ~~[40] [Smith et al 2007]~~ and a linear function of the climatic and environmental predictors. The unstructured component was represented as Gaussian distribution with zero mean. The spatiotemporal component was represented by a stationary Gaussian process[42] with covariance defined by a spatially anisotropic version of the space-time covariance function proposed by Stein (2005)[43]. To partly model seasonality, the covariance function was modified to allow the time-marginal model to include a periodic component of wavelength 12 months in the temporal covariance structure. Each survey was referenced temporally using the mid-point (in decimal years) between the recorded start and end months. For each grid location samples of the annual mean of the full posterior distribution of $Pf\text{PR}_{2\text{-}10}$ for each year were generated. These $Pf\text{PR}_{2\text{-}10}$ samples were then used to generate continuous maps of the annual mean. To determine the probable maximal malaria risk the highest value of predicted mean annual $Pf\text{PR}_{2\text{-}10}$ value at each 1 × 1 km grid location over the period 2007 to 2010 were extracted. These were then combined to generate a single map of maximum mean $Pf\text{PR}_{2\text{-}10}$.

Assessing uncertainty of model predictions

As a first step to understanding the uncertainty around the predictions of $Pf\text{PR}_{2\text{-}10}$ using the Bayesian geostatistical model, the continuous mean maps were accompanied by estimates of the posterior standard deviation. For the maximum mean $Pf\text{PR}_{2\text{-}10}$ map, the posterior standard deviations associated with the selected

mean value was used. To allow for a scaled comparison of the uncertainty of the 2010 *Pf*PR$_{2-10}$ map and the 2010 annual mean *Pf*PR$_{2-10}$, the coefficient of variation, which is a measure of dispersal around the mean[44], was computed as the ratio of the standard deviation to the mean. Higher values of the coefficient of variation suggest increasing uncertainty of model predictions. In addition, a spatially representative validation set of *Pf*PR$_{2-10}$ survey data were also selected using a spatially declustered sampling algorithm[41]. The annual predictions were then repeated in full using the remaining data to predict mean *Pf*PR$_{2-10}$ at the validation locations. The ability of the model to predict point-values of *Pf*PR at unsampled locations was quantified using two simple summary statistics: the mean prediction error (MPE); and the mean absolute prediction error (MAPE). The ME provides a measure of the model bias, while the MAE is a measure of the average accuracy of individual predictions.

Defining district level malaria endemicity and population at risk

A 2010 population surface for Somalia at 100 × 100 m spatial resolution was provided by the AfriPop project[29,30]. Using this map, estimates of the total population of each 1 × 1 km pixel to which mean *Pf*PR$_{2-10}$ was predicted was computed in ArcGIS 10. The 1 × 1 km grid squares with attached population estimates were further classified separately by the mean *Pf*PR$_{2-10}$ 2010 and the maximum mean *Pf*PR$_{2-10}$. To weight endemicity for population distribution, only those grid squares with population were retained and from these the mean *Pf*PR$_{2-10}$ was computed for each district. Based on the aggregate mean *Pf*PR$_{2-10}$ districts were then classified using a modification of the classical malaria endemicity classification[45]. The hypoendemic class (≤10% *Pf*PR$_{2-10}$) was split into <1% *Pf*PR$_{2-10}$; 1% to <5% *Pf*PR$_{2-10}$ and 5% to 10% *Pf*PR$_{2-10}$. The endemicity class of <1% *Pf*PR$_{2-10}$ represents the threshold at which an area is considered to be under low stable endemic control and a decision for sustaining control or aiming for elimination can be made[3,6]. The hyperendemic (>50% to 75% *Pf*PR$_{2-10}$) and holoendemic (>75% *Pf*PR$_{2-10}$) classes were also combined while the mesoendemic class (>10% to 50% *Pf*PR$_{2-10}$) remained unchanged. The number of districts and population by these endemicity classes were then summarised based on both the 2010 mean *Pf*PR$_{2-10}$ (contemporary risk) map and the maximum mean *Pf*PR$_{2-10}$ (receptive risk) map.

RESULTS

Predictions of mean annual PfPR$_{2-10}$ to 2010 and maximum mean posterior PfPR$_{2-10}$

A total of 1,558 *P. falciparum* community surveys (Figure 1) in which 103,593 persons were examined were assembled for the period 2007-2011 in Somalia. The majority of the data were located in the Central South zone where most of the population live. Survey data were collected across nine different months over the four years with the majority of data (76%) assembled in the months of November, December, May and June

corresponding to the peak malaria seasons in Somalia. The results of the total-set analysis showed that the model with urbanisation, precipitation, EVI and distance to main water bodies and floodplains as the best fit in predicting *Pf*PR and these variables were subsequently included in the malaria prediction model (SI 1). TSI was not selected as a statistically strong predictor of *P. falciparum* prevalence in Somalia.

The continuous 2010 malaria endemicity map for Somalia *Pf*PR_{2-10} showed the majority of locations were predicted to have parasite prevalence of <5% indicating largely hypoendemic transmission (Figure 2A). The majority of areas in North East and North West zones were predicted to be under *Pf*PR_{2-10} <12%. In contrast, the maximum annual mean *Pf*PR_{2-10} map showed a substantially different risk landscape with the majority of the Central South zone having *Pf*PR_{2-10} of >10% and an a maximum predicted mean of 38%, suggesting that peak malaria transmission in all of Central South zone and southern parts of North East zone is mesoendemic (Figure 2B). In the northern zones, maximum mean risks were predicted to be predominantly between 54% to <10% *Pf*PR_{2-10}.

The MPE and MAPE associated with the full space-time geostatistical model was 4.8% and 0.2% respectively. The 2010 annual mean *Pf*PR_{2-10} predictions were associated with higher coefficients of variation compared to the maximum mean *Pf*PR_{2-10} predictions although the difference was moderate (Figure 2C and Figure 2D). In both maps, uncertainty appeared highest in northern zones where data in space and time were fewest.

District estimates of contemporary and receptive malaria risk

According to the contemporary district malaria endemicity map based on the annual mean *Pf*PR_{2-10} map of 2010 all districts in Somalia were under hypoendemic transmission (Figure 3A). Out of the 74 districts an estimated 17 (23%) districts covering about 1.1 million people (13%) were in the <1% *Pf*PR_{2-10} risk class (Table 1 and Figure 3A). The majority of the districts (58%) and population (55%) were in the risk class of 1% to <5% *Pf*PR_{2-10} and the rest were under risks of 5% to 10% *Pf*PR_{2-10}.

In contrast the receptive risk map showed that there were no district under low stable endemic control (<1% *Pf*PR_{2-10}) and the majority of the districts (65%) and population (69%) were under the mesoendemic class (Table 1 and Figure 3B) with an upper district maximum mean estimate of 35% *Pf*PR_{2-10}. About 27% of the districts and 17% of the population were in the upper range of hypoendemicity (5% to 10% *Pf*PR_{2-10}). The rest of the districts and population were in the intermediate hypoendemic class of 1% to <5% *Pf*PR_{2-10}.

DISCUSSION

The malaria risk maps that are commonly available to countries in Africa to support malaria control planning are those that represent the contemporary distribution of risk[8, 16, 46-50]. They have been developed primarily from geo-coded parasite rate survey data[7] usually to predict risk to the most recent data year and therefore reflect transmission under scaled interventions during the era of the RBM partnership[51, 52]. In this study we argue that, in addition to contemporary maps of malaria risks, low stable endemic control and unstable transmission countries require maps of receptivity to assess the risks of rebound and epidemics and decide on where to scale-up and/or sustain intervention coverage. To demonstrate this we used community *Pf*PR survey data from the period 2007 to 2010 within a space-time MBG framework to generate two continuous malaria risk maps for Somalia. One is a contemporary map of annual mean $PfPR_{2-10}$ predicted to 2010 (Figure 2A) and the other is the maximum annual mean $PfPR_{2-10}$ map derived from the highest mean $PfPR_{2-10}$ value predicted to a location in any year over the study period to approximate receptivity (Figure 2B). We resolved these maps to the district, which is the malaria decision making unit in Somalia, and classified them by endemicity using population-weighted mean $PfPR_{2-10}$ (Figure 3).

The efficacy and impact of malaria interventions on disease in an area are dependent on its intrinsic transmission potential[53-55]. This is the theoretical basis upon which international guidelines for malaria control are formulated[1,56]. One of the most important applications of malaria risk maps for control planning, therefore, is to inform the spatial targeting of the appropriate mixes of interventions[7,57]. For Somalia, the results of the comparison of the contemporary and the receptive risk maps represents two very different transmission scenarios (Table 1; Figure 3). The contemporary malaria risk map predicted that all of Somalia was under conditions of hypoendemic transmission (≤10% $PfPR_{2-10}$) in 2010 with a fifth of the districts under risks of <1% $PfPR_{2-10}$ while the majority of the districts and population were in the intermediate hypoendemic transmission class of 1% to <5% $PfPR_{2-10}$. Under these transmission conditions targeted distribution of long lasting insecticidal nets (LLINs) and indoor residual spraying (IRS) aimed at control of residual foci are recommended while intermittent presumptive treatment in pregnancy (IPTp) is not[56]. Instead of being part of the broader monitoring and evaluation process, disease surveillance is also regarded as an intervention in of itself[1] comprising high quality passive case detection, case notification and active case detection in which all febrile cases within proximity of the index case are tested and those positive for malaria infection are radically treated[58]. In the districts where transmission is <1% $PfPR_{2-10}$, malaria elimination is considered to be technically feasible[3,6] presenting an opportunity to re-orient the sub-national strategy here towards elimination and undertake an assessment of its operational feasibility[5]. In contrast, the receptive risk map predicted that over 65% of the districts and population were under mesoendemic transmission (>10% to 50% $PfPR_{2-10}$) with the

rest exposed to hypoendemic transmission. Using this map, in the hypoendemic districts LLINs and IRS would be better targeted to foci of risk and in preparation for possible epidemics as universal coverage with these interventions is unlikely to be the most cost-effective. In those areas of receptive mesoendemic transmission, which comprise 65% of the districts, universal coverage with LLINs should be the sustained ambition[1,2,56]. The two divergent potential national malaria strategies emanating from the two different descriptions of risk highlight the danger of relying on contemporary risk maps to make decisions that require the understanding of the intrinsic transmission potential of malaria.

Available maps that describe pre-RBM distribution of risks are either expert opinion maps[59] or climate-based deterministic transmission suitability models[60] and not driven by empirical data. Even where empirical data may be available, in countries with unstable malaria transmission susceptible to seasonal and anomalous climatic variations such as the recent drought in Somalia, the risks measured at one time point may not be representative of the possible peak risk levels for that point. Therefore spatially nationally representative data over several years are required to capture these variations and estimate the highest possible transmission. In this study, *Pf*PR data for Somalia that is available over four consecutive years has provided a unique opportunity to develop a novel way of selecting the maximum predicted risks within the time series. The resolution of risk levels at the malaria resource decision-making unit also represents a product that is likely to be of more policy relevance to the national programme managers compared to the more common pixel-level predictions of risks.

The study has some limitations. Although Somalia represents one of the few African countries with ubiquitous *Pf*PR data in space and time there are gaps in the distribution of the data and uncertainty of the predictions are partly a function of these. The validation tests, however, show overall good predictive model performance with overall bias of MPE of <5% and a slight average over-prediction of about MAPE of 0.2%. The coefficient of variation, which is the ratio of standard deviation to mean *Pf*PR_{2-10}, appeared similar for both the 2010 mean and the maximum mean maps with uncertainty highest in northern zones where data in space and time were fewest (Figure 2 C & D). In selecting the maximal mean risk as predicted to a location in any year over the four year period, we make assumptions as if the modelled predictions were part of four-year repeat 'observations' of *Pf*PR_{2-10}, in that location. The basis for this is that if the mean estimate at any 1 × 1 km location from the full posterior distribution of the space-time model is a robust estimate to the given time and location, then selected maximum mean estimate is equally so. We suggest that this is a plausible assumption but further efforts need to be invested in probabilistically selecting the maximum mean predictions from the series of usually spatially and temporally uneven data. Any uncertainties in the approach used to select the maximum

annual mean *Pf*PR$_{2-10}$ are however unlikely to be the source of the major differences in endemicities when compared to the annual mean *Pf*PR$_{2-10}$ for 2010. The *Pf*PR data used in this analysis were assembled during a period when access to control interventions ~~were scaled up~~had increased in Somalia with donor funding support. Although coverage of main vector control interventions remain modest[22] it is likely that some of observed data were influenced by these interventions and in parts of the country true receptivity may even be higher than estimated. Land use changes due to urbanization, large-scale agricultural schemes and hydroelectric power dam projects also act~~s~~ as modifiers of transmission and in mapping malaria risk these factors must be adjusted for. Urbanization, which has been shown to reduce malaria transmission, was included in the analysis of receptivity for Somalia. The maps of water bodies and vegetation used in the analysis will to some degree capture any aquatic or agricultural land use changes. However, due to the long civil and the lack of a functioning central government such changes have been limited.

To compute a single estimate of risk for a relatively large area, such as districts in Somalia, will always obscure some of the heterogeneity in malaria distribution within that area regardless of the methods used. Any decision to do so is therefore a compromise between the practical applications of such a classification and the potential loss of precision in risk estimation. Approaches that directly adopt the heterogeneous properties of the prevalence data to make statistically robust single estimates of mean *Pf*PR to an administrative unit are computationally and methodologically intensive[61] but have the advantage of estimating the area level uncertainty classification through joint simulation. In this study we have used simpler approaches to partly capture the within-district heterogeneity in malaria risk when classifying them into a single endemicity class by first assigning pixel-level population to an endemicity class before aggregating to the district. Future efforts should explore approaches such as joint simulation[61] and small area estimation[62] techniques to describe the uncertainties around area level estimates of risks robustly. Such measures of uncertainties are not only quantitative estimates of model validity but also help determine where future data assembly must be prioritised to improve precision.

In conclusion, the aim of this study was to demonstrate the need for malaria receptivity maps for optimal malaria resource planning in countries which have either achieved low stable endemic control or are of unstable transmission and therefore susceptible to seasonality of climatic anomalies. We have used approaches that derive maps of contemporary malaria risk and approximations of receptivity within the same space-time MBG model resolved at the district level in Somalia. The two maps show significantly divergent transmission scenarios in which the contemporary map describes the majority of Somalia as hypoendemic and while the other shows a largely mesoendemic transmission profile. These disparities have far-reaching

consequences on decisions regarding the design and scale-up of interventions in Somalia. The results have important control implications for several low transmission countries in Africa. Urgent efforts must therefore be invested in assembling detailed historical data on parasite prevalence to allow for a better understanding of receptivity and equip national programmes with reliable estimates of receptivity that will enhance better decision making.

Acknowledgements The community survey parasite prevalence data was provided by the Food and Agricultural Organization-Food Security and Nutritional Analysis Unit (FAO-FSNAU) for Somalia. The authors are grateful ~~to~~ for the support and diligence of the all the FSNAU field staff who did a great job under difficult conditions. We are grateful to all household members who participated in the survey and agreed to the malaria testing. We thank Jacob Ouko and Damaris Kinyoki for their help with assembly of ancillary data. The authors are also grateful for comments on earlier drafts of the manuscript from Dr Emelda Okiro. This paper is published with permission from the Director of KEMRI.

Contributors AMN was responsible for overall scientific management, study design, data cleaning, analysis, interpretation, drafting and production of the final manuscript. VAA was responsible for data cleaning, geo-coding, analysis and contributed to the final manuscript. APP was responsible for the coding of the MBG models, developed the supplementary information on model specifications and contributed to final manuscript. GM and MB contributed to the survey design, data assembly and cleaning and contributed to final manuscript. FY and JA contributed to survey design, interpretation of results and contributed to final manuscript. RWS provided scientific guidance and contributed to the analysis, interpretation and preparation of the final manuscript. All authors read and approved the final manuscript.

Funding: Cross-sectional survey was funded by the FAO-FSNAU and partners. AMN is supported by the Wellcome Trust as an Intermediate Research Fellow (#095127). RWS is supported by the Wellcome Trust as Principal Research Fellow (#079080) that also funded support to APP. Programmatic support for this study was also provided through a Wellcome Trust Major Overseas Programme grant to the KEMRI/Wellcome Trust Research Programme (#092654).

Competing Interests None**.**

Ethics approval Ethical approval was provided through permission by the Ministry of Health Somalia, Transitional Federal Government of Somalia Republic, Ref: MOH/WC/XA/146./07, dated 02/02/07. Informed verbal consent was sought from all participating households and individuals. Participants who were positive for *Plasmodium falciparum* infection was treated with the correct dosages of the nationally recommended antimalarial.

Data sharing statement Data from this study are not in the public domain

References

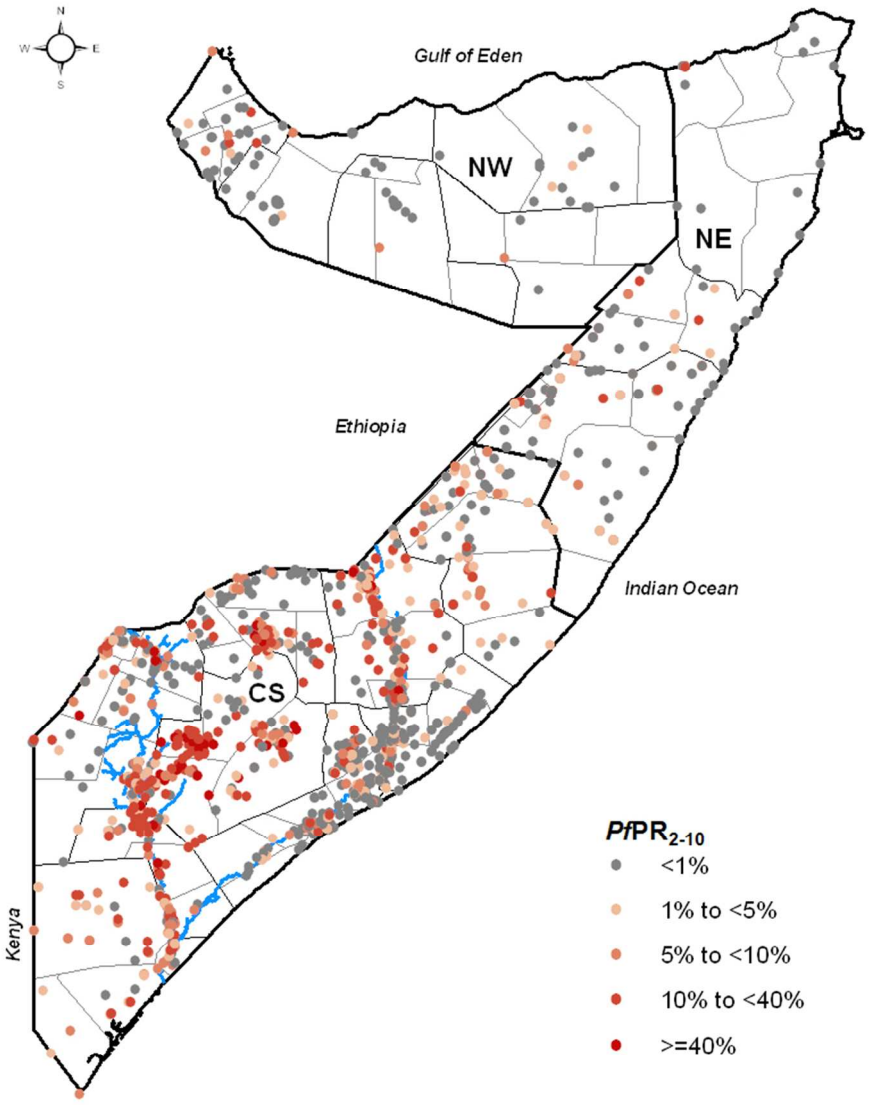
1. World Health Organization: Malaria elimination: a field manual for low and moderate endemic countries, Geneva: World Health Organization; 2007.
2. Feachem RGA, Phillips AA, Hwang J, *et al*. Shrinking the malaria map: progress and prospects. *Lancet* 2010; 376: 1566 – 1578.
3. Cohen JM, Moonen B, Snow RW, et al. How absolute is zero? An evaluation of historical and current definitions of malaria elimination. *Malar J* 2010; 9: 213.
4. Tatem AJ, Qiu Y, Smith DL, et al. The use of mobile phone data for the estimation of the travel patterns and imported Plasmodium falciparum rates among Zanzibar residents. *Malaria J* 2009; 8: 287.
5. Tatem AJ, Smith DL. International population movements and regional *Plasmodium falciparum* malaria elimination strategies. *Proceedings of the National Academy of Sciences of the United States of America*, 2010; 107: 12222 - 12227.
6. Snow RW, Marsh K. Malaria in Africa: progress and prospects in the decade since the Abuja Declaration. *Lancet* 2010; 376: 137-139.
7. Hay SI, Snow RW. The Malaria Atlas Project: Developing global maps of malaria risk. *PLoS Med* 2006; 3: e473.
8. Gething PW, Patil AP, Smith DL, *et al*. A new world malaria map: Plasmodium falciparum endemicity in 2010. *Malar J* 2011; 10: 378 doi: 10.1186/1475-2875-10-378.
9. Snow RW, Amratia P, Kabaria CW, et al. The changing limits and incidence of malaria in Africa: 1939-2009. *Adv Parasitol* 2012; in press.
10. Hadden RL. *The Geology of Somalia: A Selected Bibliography of Somalian Geology, Geography and Earth Science*. Engineer Research and Development Laboratories, Topographic Engineering Centre. http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA464006&Location=U2&doc=GetTRDoc.pdf. Accessed 12 December 2011.
11. FAO-SWALIM Databases. http://geonetwork.faoswalim.org/geonetwork/srv/en/main.home. Accessed 10th December 2011.
12. Choumara R. Notes sur le paludisme au Somaliland. *Riv di Malariol* 1961; 40: 9-34.
13. Kamal M. *Entomological surveillance in Somalia*. In Consultancy report for WHO Somalia, 2007.
14. Ilardi I, Sebastian A, Leone F, *et al*. Epidemiological study of parasitic infections in Somali nomads. *Trans R Soc Trop Med Hyg* 1987; 81: 771-772.
15. Warsame M, Perlmann H, Ali S, *et al.* The sero-reactivity against Pf155 (RESA) antigen in villagers from a meso-endemic area in Somalia. *Trop Med & Parasitol* 1989; 40: 412-414.
16. Noor AM, Clements ACA, Gething PW, *et al.* Spatial prediction of *Plasmodium falciparum* prevalence in Somalia. *Malar J* 2008; 7: 159.
17. Bousema T, Jackie C, Patrick C, *et al*. Serological markers for low exposure to malaria. *Emerg Infect Dis* 2009; 16: 392-399.
18. The Crisis Group. Somalia. http://www.crisisgroup.org/en/regions/africa/horn-of-africa/somalia.aspx, Accessed 12 November 2011.
19. The Transitional Federal Government of Somalia http://en.wikipedia.org/wiki/History_of_the_Transitional_Federal_Government_of_the_Republic_of_Somalia, Accessed 12 December 2011.
20. UNOPS 2010. The Somali Support Secretariat Project Board Terms of Reference as promulgated at the 10 May 2010 Project Board Meeting. http://www.coordinate4somalis.info/images/stories/pdf/Somali%20Support%20Secretariat%20Project%20Board%20Terms%20of%20Reference.pdf. Accessed 7 March 2012.
21. GFATM Project proposal.http://www.theglobalfund.org/search/docs/6SOMM_1418_0_full.pdf. Accessed 10 December 2007.
22. World Health Organization-Roll Back Malaria. Somalia National Strategic Plan for Malaria 2011 – 2015; 2010.
23. World Health Organization-Roll Back Malaria. National Malaria Prevention and Control Monitoring and Evaluation Plan 2011 – 2015; 2010.
24. GFATM Progress report. http://www.theglobalfund.org/en/library/publications/progressreports/. Accessed 15 December 2011.
25. Smith DL, McKenzie FE, Snow RW, et al. Revisiting the basic reproductive number for malaria and its implications for malaria control. *PLoS Biol* 2007; 5: e42.

26. Smith DL, Dushoff J, Snow RW, et al. The entomological inoculation rate and *Plasmodium falciparum* infection in African children. *Nature* 2005; 438: 492-495.
27. FSNAU Analytical systems. http://www.fsausomali.org/200510112504_baselines.php. Accessed 11th January 2012.
28. Molineaux L, Muir DA, Spencer HC, et al. *The epidemiology of malaria and its measurement*. Malaria: Principles and Practice of Malariology. W. H. Wernsdorfer and I. McGregor. London, Churchill Livingstone, 1988; 2: 999-1089
29. Linard C, Alegana VA, Noor AM, et al. A high resolution spatial population database of Somalia for disease burden estimation. *Intl J of Health Geogr* 2010; 9: 45.
30. The AfriPop Project. http://www.clas.ufl.edu/users/atatem/index_files/Details.htm, Accessed 15 January 2011.
31. Hijmans RJ, Cameron SE, Parra JL, et al.Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol* 2005; 25: 1965-1978.
32. WorldClim-Global Climate Database. www.wordlclim.org. Accessed 10 February 2012.
33. MODIS- EVI data archives. ftp://n4ftl01u.ecs.nasa.gov/SAN/MOST/MOD10A2.005/. Accessed 12 September 2011.
34. Global Lakes and Wetlands Database Request. https://secure.worldwildlife.org/science/data/item1877.html. Accessed 10 September 2011.
35. Gething PW, Van Boeckel T, Smith DL, *et al*. Modelling the global constraints of temperature on transmission of *Plasmodium falciparum* and *P. vivax*. *Parasite Vector* 2011; 4:92.
36. Miller A. Subset Selection in Regression. Boca Raton, FL: Chapman & Hall. 2002:238 p.
37. Lumley T leaps: regression subset selection (R package) Version 2.7, 2010.
38. Schwarz G. Estimating dimensions of a model. *Ann Stat* 1978; 6: 461-464.
39. Diggle PJ. Spatio-temporal point processes, partial likelihood, foot and mouth disease. *Stat Methods Med Res* 2006; 15:325-36.
~~39. Diggle P, Moyeed R, Rowlingson B, Thompson M. Childhood malaria in the Gambia: a case-study in model-based geostatistics. *Appl Stat* 2002; 51: 493-506.~~
40. Smith DL, Guerra CA, Snow RW, et al. Standardizing estimates of the Plasmodium falciparum parasite rate. *Malar J* 2007; 6: 131.
41. Malaria Atlas Project *P. falciparum* Cartographic code. https://github.com/malaria-atlas-project/mbgw-clean. Accessed October 2011.
42. Banerjee S, Carlin BP, Gelfand AE. *Hierarchical modeling and analysis for spatial data. Monographs on Statistics and Applied Probability 101*. Boca Raton, Florida, U.S.A.: Chapman & Hall / CRC Press LLC, 2004.
43. Stein ML. Space-time covariance functions. *J Am Stat Assoc* 2005; 100: 310-321.
44. Kirkwood TBL. Geometric means and measures of dispersion. *Biometrics* 1979; 35:908-909.
45. Metselaar D, PH van Thiel. Classification of malaria. *Trop Geogr Med* 1959; 11: 157-161.
46. Noor AM, Gething PW, Alegana VA, *et al*. The risks of malaria infection in Kenya in BMC *Infect Dis* 2009; 9*:* e180.
47. Gosoniu L, Veta AM, Vounatsou P. Bayesian Geostatistical Modeling of Malaria Indicator Survey Data in Angola. *PLoS ONE* 2010*;* 5: e9322. doi:10.1371/journal.pone.0009322.
48. Riedel N, Vounatsou P, Miller JM, *et al*. Geographical patterns and predictors of malaria risk in Zambia: Bayesian geostatistical modelling of the 2006 Zambia national malaria indicator survey (ZMIS). *Malar J* 2010; 9: 37.
49. Taylor SM, Messina JP, Hand CC, *et al*. Molecular Malaria Epidemiology: Mapping and Burden Estimates for the Democratic Republic of the Congo, 2007. *PLoS ONE* 2011; 6(1): e16420. doi:10.1371/journal.pone.0016420
50. Stensgaard A-S, Vounatsou P, Onapa AW, *et al*. Bayesian geostatistical modelling of malaria and lymphatic filariasis infections in Uganda: predictors of risk and geographical patterns of co-endemicity. *Malar J* 2011; 10:298
51. Roll Back Malaria. Progress and Impact Series Number 2. http://www.rbm.who.int/ProgressImpactSeries/docs/RBMMalariaFinancingReport-en.pdf, Accessed 16 February 2012.
52. Roll Back Malaria. Progress and Impact Series Number 2. http://www.rbm.who.int/ProgressImpactSeries/docs/wmd2010report-en.pdf, Accessed 16 February 2012.
53. Snow RW, Marsh K. The consequences of reducing transmission of Plasmodium falciparum in Africa. *Adv Parasitol* 2002; 52: 235-264.
54. Killeen GF, Smith TA, Ferguson HM, *et al.* Preventing Childhood Malaria in Africa by Protecting Adults from Mosquitoes with Insecticide-Treated Nets. *PLoS Med* 2007; 4: e229.
55. Smith DL, Noor AM, Hay SI, et al. Predicting changing malaria risk following expanded insecticide treated net coverage in Africa. *Trends Parasitol* 2009, 25: 511-516.
56. World Health Organization. Manual for developing a national malaria strategic plan, Geneva, WHO, September 2011.

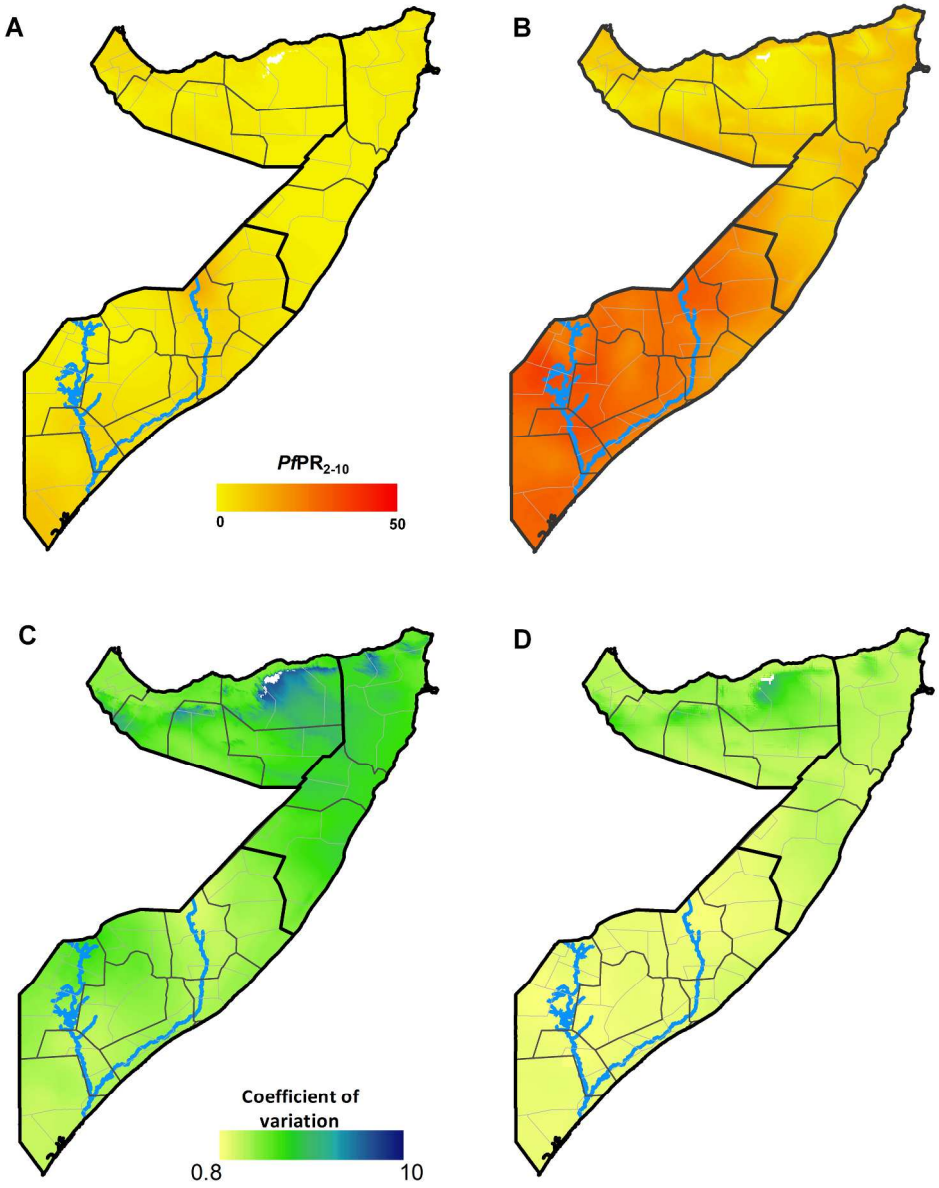
57. Noor AM, Alegana VA, Patil AP, et al. Predicting the unmet need for biologically targeted coverage of insecticide treated nets in Kenya. *Am J Trop Med Hyg* 2010*,* 83: 854-860.
58. Moonen B, Cohen JM, Snow RW, *et al*. Operational strategies to achieve and maintain malaria elimination. *Lancet* 2010; 376: 1592 – 1603.
59. Lysenko AJ, Semashko IN: Geography of malaria. A medico-geographic profile of an ancient disease [in Russian]. In: *Itogi Nauki: Medicinskaja Geografija.* Edited by Lebedew AW.Moscow: *Acad. Sci*, USSR; 1968: 25-146.
60. Craig MH, Snow RW, le Sueur D. A climate-based distribution model of malaria transmission in sub-Saharan Africa. *Parasitol Today* 1999; 15: 105–111.
61. Gething PW, Patil AP, Hay SI. Quantifying aggregated uncertainty in *Plasmodium falciparum* malaria prevalence and populations at risk via efficient space-time geostatistical joint simulation. *PLoS Comput Biol* 2010*;* 6: e1000724.
62. Bayesian methods for combining multiple Individual and Aggregate data Sources in observational studies. http://www.bias-project.org.uk/software/#sae. Accessed 15 February 2011.

Table 1 A summary of districts (N=74) and population in 2010 (N=8.4 million) in Somalia classified by malaria endemicity. District classifications of endemicity were computed from population-weighted posterior annual mean *Pf*PR_{2-10} predicted to 2010 (contemporary) and the maximum annual mean *Pf*PR_{2-10} (receptive) predictions over the period 2007-2010.

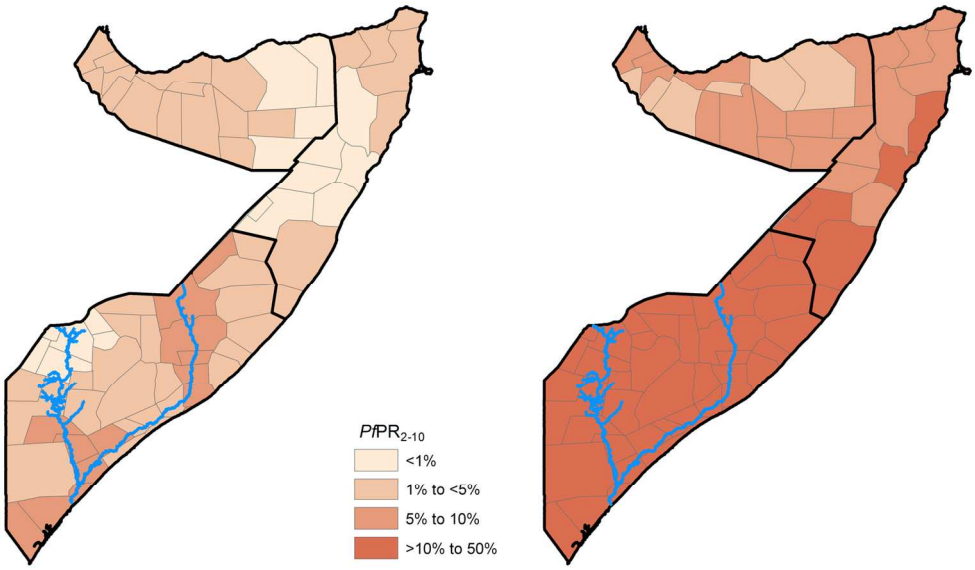
	Endemicity classification based on the 2010 annual mean *Pf*PR_{2-10} (contemporary) predictions		**Endemicity classification based on the maximum mean *Pf*PR_{2-10} (receptive) predictions over the period 2007-2010**	
	Number (%) of districts	**Population at risk, million, (%)**	**Number (%) of districts**	**Population at risk, million, (%)**
Population weighted mean *Pf*PR_{2-10}				
Hypoendemic				
<1% *Pf*PR_{2-10}	17 (23)	1.1 (13)	0 (0)	0 (0)
1% to <5% *Pf*PR_{2-10}	43 (58)	4.6 (55)	6 (8)	1.2 (14)
5% to 10% *Pf*PR_{2-10}	14 (19)	2.6 (31)	20 (27)	1.4 (17)
Mesoendemic (>10% to 50% *Pf*PR_{2-10})	0 (0)	0.0 (0)	48 (65)	5.8 (69)
Hyperendemic and Holoendemic (>50% *Pf*PR_{2-10})	0 (0)	0.0 (0)	0 (0)	0.0 (0)



Zone, regional and district maps of Somalia showing the distribution of the age-standardised community Plasmodium falciparum parasite rate (PfPR2-10) data (n=1,558) assembled during the period 2007 – 2010 (including 54 surveys undertaken in 2011). The zones are CS= Central South; NE= North East; NW= North West. The thick black line show the zone boundaries, the thin black lines show the regional boundaries and the thin grey lines show the district boundaries. The blues lines show the location of the Juba (lower) and Shabelle (upper) Rivers.
215x279mm (96 x 96 DPI)



A) Map of the posterior annual mean PfPR2-10 prediction to 2010 (contemporary) at 1×1 km grid location in Somalia; B) Map of the maximum mean PfPR2-10 prediction (receptive) at 1×1 km grid location as computed from the posterior annual mean PfPR2-10 prediction for each year from 2007 to 2010; C) Map of the coefficient of variation (the standard deviation/the mean PfPR2-10 prediction) of the contemporary prediction at 1x1 km grid location; D) Map of the coefficient of variation at 1x1 km grid location of the receptive prediction. The thick black lines show the zone boundaries, the thin black lines show the regional boundaries and the thin grey lines show the district boundaries. Higher coefficient of variation of the predictions suggests higher uncertainties of the PfPR2-10 predictions. The scale bar for the continuous PfPR2-10 ends at 50 which is the upper limit of mesoendemic transmission. The blues lines show the location of the Juba (lower) and Shabelle (upper) Rivers.
253x321mm (300 x 300 DPI)



District (n=74) maps of Somalia classified by endemicity using a population-weighted: A) posterior aggregate annual mean PfPR2-10 (contemporary) prediction to 2010 and; B) the maximum annual mean PfPR2-10 (receptive) predictions over the period 2007-2010. The blues lines show the location of the Juba (lower) and Shabelle (upper) Rivers.
148x89mm (300 x 300 DPI)

Supplementary Information (SI) 1: Predictors of age-standardised *Plasmodium falciparum* parasite rate ($Pf PR_{2-10}$)

SI 1.1 Urbanisation: A surface of urbanisation in Somalia derived from a 100×100 m spatial resolution population surface developed from a combination of census, satellite imagery and land cover data[1,2] was used to define each survey location as urban or rural. Somalia population map was derived by a combination of simple and semi-automated processes involving the re-distribution of district level population estimates to a finer spatial scale using land cover and land use datasets[1]. A refined land cover map was prepared by combining the settlement data with Africover data[3] and other ancillary GIS data on roads, rivers from Vector Map Level Zero (VMAP 0)[4] and the Landsat satellite imagery. A total of 22 land cover classes were formed and population density within these classes was calculated with urban and rural extents derived from the GEOterrain counsultancy[5]. These calculated densities were then used as weightings to redistribute population by settlement and in land cover types that were unaccounted for by existing settlement size data (Figure SI 1A).

SI 1.2 Enhanced vegetation Index (EVI): EVI is an index of intensity of photosynthetic activity and a good proxy for rainfall ranges from 0 (no vegetation) to 1 (complete vegetation). Monthly EVI surfaces at 1×1 km spatial resolution derived from the global Moderate Resolution Imaging Spectro-radiometer (MODIS) satellite imagery for the period 2001-2010 were downloaded from the MODIS webiste[6] and were used to compute annual mean EVI. These monthly maps were used to compute annual mean EVI for the years 2007 to 2010 (Figure SI 1B).

SI1.3 Precipitation: Monthly mean precipitation (mm) raster surfaces at 1×1 km resolution were downloaded from the WorldClim website[7] and used as a proxy for rainfall. The Worldclim database was compiled using weather stations data collected world-wide for the period 1950-2000 and interpolated using spline methods[8]. The monthly mean precipitation was used to compute annual mean precipitation surfaces for the years 2007 to 2010 (Figure SI 1C).

SI1.4 Temperature suitability Index (TSI): As metric for the effect of temperature on malaria transmission, TSI at a spatial resolution of 1×1 km[9] was used. TSI was constructed using monthly temperature time series[7] within a biological modelling framework to quantify the effect of ambient temperature on sporogony and vector survivorship and determine the suitability of an area to support transmission globally[9]. On a scale of increasing transmission suitability, TSI ranges from 0 (unsuitable) to 1 (most suitable) (Figure SI 1.1.1D).

SI 1.5 Potential mosquito breeding sites: To define potential breeding sites the Global Lakes and Wetlands databases (GLWD) developed as a partnership between the World Wildlife Fund and the Center for Environmental Systems Research, University of Kassel, Germany[10] was used. The GLWD data contained 12 wetland classifications but for Somalia the following wetland classes were retained for analysis: lake, rivers, floodplains, reservoirs and coastal wetlands. Intermittent water bodies and brackish and saline water were removed. Euclidean distances to these wetlands were computed in ArcGIS 10 (ESRI Inc. NY, USA) (Figure SI 1E).

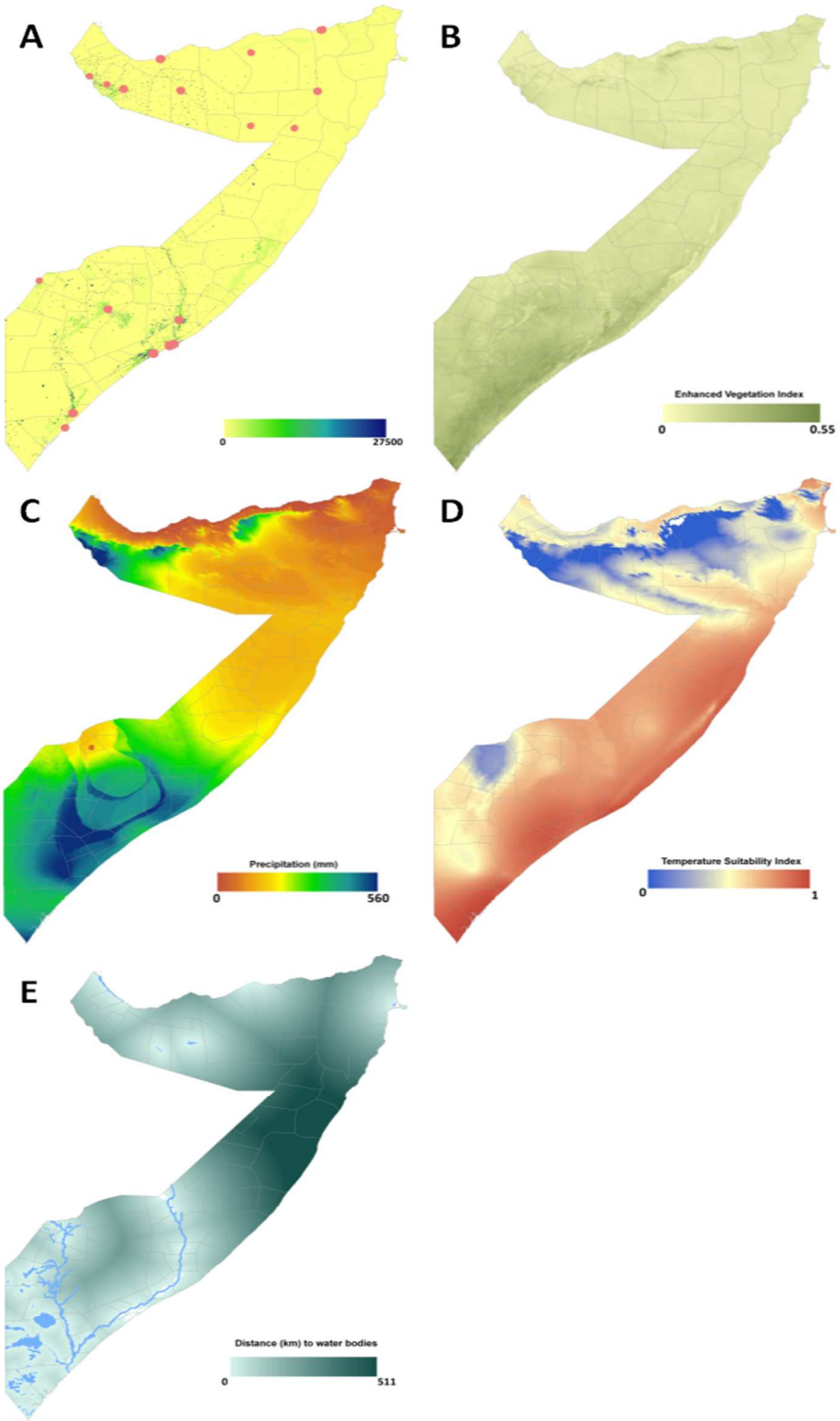
SI 1.6 Statistical analysis of the predictors of $PfPR_{2-10}$

The values of the underlying ecological, climatic and population covariates describe above were extracted to each survey location using ArcGIS 10 *Spatial Analyst* tool. Distance to potential breeding sites was log-transformed before analysis because of its high positive skew. The covariates were then included in total-sets analysis which is an automatic model selection process based on a generalized linear regression model and implemented using the *bestglm* package in R[11,12]. This approach selects the best combination of the covariates based on the value of the Bayesian Information Criteria (BIC) statistic[13] which selects the lowest BIC as the best model fit. The model excluding TSI showed the lowest BIC and therefore the best fit (Table SI 1.1.1).

Table SI 1.1.1 Summary of regression analysis of covariates

Covariate	Odds Ratio (95% CI)	P-value
Urbanisation	-0.88 (-1.02- -0.69)	<0.001
EVI	0.81 (0.19-1.44)	0.011
Precipitation	0.003 (0.002-0.004)	<0.001
TSI	-0.68 (-0.89- 0.01)	<0.312
Log of distance to wetlands	0.27 (0.24 – 0.31)	<0.001

Figure SI 1 Somalia maps of 1×1 km spatial resolution of: **A)** population distribution showing the location of urban centres (in red); **B)** annual mean enhanced vegetation index (EVI); **C)** annual mean precipitation (mm); **D)** temperature suitability index for *Plasmodium falciparum* transmission; **E)** distance (km) to nearest water body



References

1. Linard C, Alegana VA, Noor AM, Snow RW *et al.,* A high resolution spatial population database of Somalia for disease burden estimation. *International Journal of Health Geographics* 2010; 9: 45.
2. The AfriPop Project. http://www.clas.ufl.edu/users/atatem/index_files/Details.htm. Accessed 15 January 2012.
3. Africover. www.africover.org. Accessed 13 February 2012.
4. VMAP 0. http://www.mapability.com/index1.html?http&&&www.mapability.com/info/vmap0_index.html. Accessed 3 January 2012.
5. GeoTerraImage. www.geoterraimage.com. Accessed 15 February 2012.
6. Nasa TRMM. Goddard, Maryland, USA, NASA Goddard Space Flight Center. http://trmm.gsfc.nasa.gov/. Accessed 15 November 2011.
7. WorldClim. http://www.worldclim.org/download.htm. Accessed 10 December 2011.
8. Hijmans RJ, Cameron SE, Parra JL, Jones PG *et al.*.Very high resolution interpolated climate surfaces for global land areas. *Int J Climatology* 2005; 25: 1965-1978.
9. Gething PW, Van Boeckel T, Smith DL, *et al.* Modelling the global constraints of temperature on transmission of *Plasmodium falciparum* and *P. vivax*. *Parasite Vector* 2011; 4:92.
10. Global Lakes and Wetlands Database Request. https://secure.worldwildlife.org/science/data/item1877.html. Accessed 10 September 2011.
11. Miller A. *Subset Selection in Regression.* Boca Raton, FL: Chapman & Hall. 2002:238 p.
12. Lumley T. *leaps: regression subset selection* (R package) Version 2.7, 2010.
13. Schwarz G. Estimating dimensions of a model. *Ann Stat* 1978; 6: 461-464.

Supplementary Information (SI) 2: Model-based Geostatistical Procedures

Below are the details of the MBG framework used to develop the contemporary and receptive malaria risk maps for Somalia. The MBG procedure was adopted from Gething et al 2011[1] where further model details, especially on age-standardisation of parasite rate data, can also be found. The generic model code is available on Malaria Atlas Project *P. falciparum* Cartographic code link[2] and has been adapted for Somalia. Model fitting was achieved using Markov chain Monte Carlo (MCMC)[3,4].

SI 2.1 The MBG presentation

Each of the N_i individuals in sample i was assumed *P. falciparum* positive with probability $\tilde{k}_i P'(x_i, t_i)$, so the number positive N_i^+ was distributed binomially:

$$N_i^+ | N_i, P'(x_i, t_i) \overset{\text{ind}}{\sim} \text{Bin}(N_i, \tilde{k}_i P'(x_i, t_i)) \qquad \text{S2.1}$$

The coefficient $P'(x_i, t_i)$ was modelled as a Gaussian process. The factor $\tilde{k}_i$ converted $P'(x_i, t_i)$ to the probability that individuals within the age range reported for study i were *P. falciparum* positive, and that the infection was detected, thereby accounting for the influence of age on the probability of detection[5]. The age-standardisation factor $\tilde{k}_i$ in each population was assumed drawn independently from a distribution $D_{\tilde{k}}$ whose parameters were the lower $A_{L,i}$ and upper $A_{U,i}$ ages reported in study i:

$$\tilde{k}_i | A_{U,i}, A_{L,i} \overset{\text{ind}}{\sim} D_{\tilde{k}}(A_{U,i}, A_{L,i}) \qquad \text{S2.2}$$

The form of $D_{\tilde{k}}$ is described in Gething et al 2011[2].

*Pf*PR$_{2\text{-}10}$ is the *P. falciparum* parasite rate for individuals between ages 2 (2.00) and 10 (9.99). Its value at an arbitrary location x and time t is the product of $P'(x,t)$ and another age-standardisation factor, k_{2-10}, distributed as $D_k(2, 10)$:

$$\begin{aligned} &\text{PR}_{2-10}(x,t) = P'(x,t) k_{2-10}(x,t) \\ &k_{2-10}(x,t) \overset{\text{ind}}{\sim} D_k(2,10) \end{aligned} \qquad \text{S2.3}$$

The factor k_{2-10} converted $P'(x,t)$ to the probability that individuals between ages 2 and 10 at location x are *P. falciparum* positive. The age-standardisation factor $\tilde{k}$ of a survey is the product of the age-standardisation factor k associated with the same place, time and age range and the sensitivity of the survey.

The coefficient $P'(x,t)$ at arbitrary location x and time t was modelled as the inverse-logit function applied to a random field f evaluated at (x,t), plus an unstructured (random) component $\epsilon(x,t)$.

$$P'(x,t) = \text{logit}^{-1}(f(x,t) + \epsilon(x,t)) \quad \text{S2.4}$$

The components $\epsilon(x,t)$ were assumed independent and identically distributed for each location x and time t and a standard diffuse but proper prior with expectation 0.25 was assigned to their variance V.

$$\epsilon(x,t)|V \overset{\text{iid}}{\sim} \text{N}(0,V) \quad \text{S2.5}$$

$$\frac{1}{V} \sim Gamma(3,12) \quad \text{S2.6}$$

The random field f was modelled as a Gaussian process characterised by its mean and covariance functions:

$$f(x,t)|\boldsymbol{\beta},\tau,\phi_x,\phi_t,\lambda,\psi,\rho,\upsilon \sim \text{GP}(\boldsymbol{\beta},C) \quad \text{S2.7}$$

The mean function was defined as $\mu = \boldsymbol{\beta}\mathbf{X}$,where $\mathbf{X} = 1, X_1(x), ..., X_n(x)$ was a vector consisting of a constant and $n = 20$ environmental covariates indexed by spatial location x ,and $\boldsymbol{\beta} = \beta_0, \beta_1, ..., \beta_n$ was a corresponding vector of regression coefficients. The covariance of the field was modelled using a version of the spatiotemporal covariance function recently recommended by Stein[6] (equation 2.12):

$$C(x_i,t_i;x_j,t_j) = \tau^2\gamma(0)\frac{(\Delta x)^{\gamma(\Delta t)}K_{\gamma(\Delta t)}(\Delta x)}{2^{\gamma(\Delta t)-1}\Gamma(\gamma(\Delta t)+1)},$$

$$\gamma(\Delta t) = \frac{1}{2\rho+2(1-\rho)\left[(1-\upsilon)e^{-|\Delta t|/\phi_t}+\upsilon\cos(2\pi\Delta t)\right]}, \quad \text{S2.8}$$

$$\Delta t = |t_i - t_j|$$

K_γ is the modified Bessel function of the second kind of order γ, and Γ is the gamma function [7,8].

Spatial distance between a pair of points x_i and x_j was computed as great-circle distance $D_{GC}(x_i,x_j)$ multiplied by a factor that depends on the angle of inclination $\theta(x_i,x_j)$ of the vector pointing from x_i to x_j. θ was computed as if latitude and longitude were Euclidean coordinates (on a cylindrical projection) to allow for anisotropy:

$$\Delta x = 2\sqrt{\gamma(\Delta t)}\frac{D_{GC}(x_i, x_j)\sqrt{1 - \psi^2 \cos^2(\theta(x_i, x_j) - \lambda)}}{\phi_x} \quad \text{S2.9}$$

When $\Delta x = 0$ (that is, for points at the same location but different times), the covariance function reduces to

$$\rho + (1 - \rho)\left[(1 - \upsilon)e^{-|\Delta t|/\phi_t} + \upsilon \cos(2\pi\Delta t)\right] \quad \text{S2.10}$$

As temporal separation increases, the covariance approaches a limiting sinusoid $\tau^2[\rho + (1 - \rho)\upsilon \cos(2\pi\Delta t)]$ rather than zero. When $\Delta t = 0$, on the other hand (for points at different locations but the same time), it reduces to a standard exponential form with range parameter $\phi_x\sqrt{2}$. Unlike standard sum-product models, this covariance function does not have problematic ridges along its axes[6].

SI 2.2 Prior Specification

The square root of the partial sill τ and the spatial range parameter ϕ_x were assigned skew-normal priors:

$$\log \tau | \mu_\tau, V_\tau, \alpha_\tau \sim \text{Skew-Normal}(\mu_\tau, V_\tau, \alpha_\tau) \quad \text{S2.11}$$

$$\log \phi_x | \mu_\phi, V_\phi, \alpha_\phi \sim \text{Skew-Normal}(\mu_\phi, V_\phi, \alpha_\phi) \quad \text{S2.12}$$

and their specification is described further below.

The standard "one-over-*x*" prior for the temporal scale parameter ϕ_t resulted in collapse to zero, a common artefact when data do not contain strong information. A relatively vague but proper prior, which has an expectation of ten years, was used instead.

$$\phi_t \sim \text{Exponential}(0, .1) \quad \text{S2.13}$$

A uniform prior was assigned to the direction of anisotropy parameter λ and to the square of the "eccentricity" parameter ψ, which controls the amount of anisotropy,

$$\lambda \sim \text{Uniform}(0, \pi) \quad \text{S2.14}$$

$$\psi^2 \sim \text{Uniform}(0, 1) \quad \text{S2.15}$$

a uniform prior was assigned to the temporal parameters governing the amplitude of the sinusoidal component ρ and the limiting autocorrelation in the temporal direction υ:

$$\rho \sim \text{Uniform}(0,1), \upsilon \sim \text{Uniform}(0,1) \qquad \text{S2.16}$$

and a standard prior was assigned to the components of the mean:

$$p(\boldsymbol{\beta}) \propto 1 \qquad \text{S2.17}$$

Although standard priors such as the improper "flat" prior[3] were assigned to most of the basic model parameters, subjective skew-normal priors[7] were specified for the range and partial sill parameters τ and ϕ_x.

SI 2.3 Model implementation

SI 2.3.1 MCMC Algorithms

Both the main geostatistical model and the age-standardisation sub-model were fitted using the MCMC algorithm[3,4]. The algorithm was implemented in the Python[8] and FORTRAN programming languages using the open-source Bayesian statistics package PyMC[9,10] and the numerical packages SciPy and NumPy[11].

The evaluation of f at the sampling locations and times was updated using Gibbs steps[3]. The evaluation of the uncorrelated process ϵ was updated one point at a time using random-walk Metropolis steps[3]. The model parameters $\boldsymbol{\beta}$, τ, ϕ_x, ϕ_t, λ, ψ, V and ρ were updated jointly using the method of Haario, Saksman and Tamminen[12].

Within the MCMC loop, the age-standardisation factors $\tilde{k}_i$ were not imputed explicitly. We were not interested in their particular values, and marginalizing out "nuisance parameters" ahead of time usually improves the mixing of MCMC algorithms. Before the MCMC loop began, the marginal likelihood:

$$\int \text{Bin}(N_i^+; N_i, k_i P'(x_i,t_i)) D_{\tilde{k}}(\tilde{k}_i; A_{U,i}, A_{L,i}) d\tilde{k}_i \qquad \text{S2.18}$$

was approximated using standard Monte Carlo integration for several values of $P'(x_i,t_i)$. That is, values for the model parameters α_i, b_i, c_i and s_i and the age distribution S_i were drawn from their posterior predictive distributions, then expression (S3.1) was evaluated to obtain k_i, then the binomial probability

was evaluated for several values of $P'(x_i, t_i)$. The probabilities resulting from many such draws were averaged. Inside the MCMC loop, the marginal likelihood function for arbitrary values of $P'(x_i, t_i)$ was evaluated by interpolation.

SI 2.3.2 Age Correction Model

The age distribution parameters S_i, S_0 and ν are independent of the relative *Pf*PR parameters P_i', α_i, c_i, b_i, s_i, μ_A, σ and R given the data, so these two groups of parameters were inferred using separate MCMC algorithms.

In the MCMC for the age distribution parameters, the survey populations' age distributions S_i were updated using Gibbs steps[3]. The concentration parameter ν was updated using random-walk Metropolis steps[3]. The typical age distribution S_0 was represented as a normalized sequence of gamma random variables[13], and these variables were updated one at a time using random-walk Metropolis steps[3].

In the MCMC for the relative *Pf*PR parameters, the distributional parameters μ_A, σ and R were updated jointly using the method of Haario, Saksman and Tamminen[12]. The parameters P_i', α_i, c_i, b_i and s_i were updated jointly for each population i using the same method.

SI 2.3.3 Spatiotemporal Prediction

The output of the MCMC stage consisted of $\{\theta_{(l)}; l = 1, ..., m\}$ samples from the posterior of the parameter set $\theta = \{\boldsymbol{\beta}, \tau, \phi_x, \phi_t, \lambda, \psi, \rho, k, V\}$ and a corresponding $\{f(x_i, t_i)_{(l)}; l = 1, ..., m\}$ samples from the posterior of the space-time random field at each of the n data locations$\{(x_i, t_i); i = 1, ..., n\}$. For every $l'th$ sample, the conditional distribution of the annual mean of the space-time random field was predicted at each prediction location x_j on the nodes of a regular 1×1 km grid within the spatial limits of stable *P. falciparum* transmission[14]. The distribution of the annual mean $f(x_j)_{(l)}$ for prediction location x_j was modelled as the joint multivariate normal distribution of the 12 predicted monthly values $\{t = 2007_{Jan}, ..., 2007_{Dec}\}$ for that year specified by a 12 element mean vector $\hat{\boldsymbol{y}}(x_j)_{(l)}$ and 12 × 12 variance-covariance matrix $\hat{\boldsymbol{\sigma}}^2(x_j)_{(l)}$:

$$f(x_j)_{(l)} \sim MVN\left(\hat{\boldsymbol{y}}(x_j)_{(l)}, \hat{\boldsymbol{\sigma}}^2(x_j)_{(l)}\right) \quad \text{S2.19}$$

The mean vector $\hat{\boldsymbol{y}}(x_j)_{(l)}$ was computed using:

$$\hat{\boldsymbol{y}}(x_j)_{(l)} = \boldsymbol{\mu}_{P_{(l)}} + \boldsymbol{C}^T_{DP_{(l)}} \cdot \boldsymbol{C}^{-1}_{DD_{(l)}} \cdot (\boldsymbol{p}(x,t) - \boldsymbol{\mu}_{D_{(l)}}) \quad \text{S2.20}$$

where $\boldsymbol{\mu}_P$ and $\boldsymbol{\mu}_D$ were the predicted mean of the random field at each of the 12 prediction times $\{t = 2007_{Jan}, ..., 2007_{Dec}\}$ at spatial location x_j and at each of the n data locations respectively, $\boldsymbol{C}_{DP}$ and $\boldsymbol{C}_{DD}$ were the data-to-prediction and data-to-data covariance matrices respectively, and $\boldsymbol{p}(x,t)$ was the vector of n data values. The 12 × 12 variance-covariance matrix $\hat{\boldsymbol{\sigma}}^2(x_j)_{(l)}$ was computed using:

$$\hat{\boldsymbol{\sigma}}^2(x_j)_{(l)} = \boldsymbol{C}_{PP_{(l)}} - \boldsymbol{C}^T_{DP_{(l)}} \cdot \boldsymbol{C}^{-1}_{DD_{(l)}} \cdot \boldsymbol{C}_{DP_{(l)}} \quad \text{S.2.21}$$

The value of the $l'th$ sample of V, the variance of the unstructured component $\epsilon(x,t)$, was then added to the diagonal of the matrix $\hat{\boldsymbol{\sigma}}^2(x_j)_{(l)}$ and 1000 draws were made randomly from the distribution specified in equation S2.34. These draws represented samples from the posterior distribution of $f(x_j)$ and were subject to an inverse logit transform and then multiplied by the $l'th$ sample of the age-standardisation parameter $k_{2-10(l)}$ to form the $l'th$ sample from the posterior distribution of the predicted mean annual 2010 *Pf*PR$_{2\text{-}10}$ endemicity surface at location x_j:

$$P'_{2-10}(x_j)_{(l)} = \text{logit}^{-1}\left(f(x_j)_{(l)} + \epsilon_{(l)}\right) k_{2-10(l)} \quad \text{S.2.22}$$

This procedure was repeated for every $l'th$ sample to form the set $\{P'_{2-10}(x_j)_{(l)}; l = 1, ..., m\}$ of m samples for each prediction location. The point estimate of *Pf*PR$_{2\text{-}10}$ endemicity at each location was defined as the mean of this set, whilst the probability of membership to each class was computed as the proportion of these samples falling within each class definition.

References

1. Gething PW, Patil AP, Smith DL, *et al*. A new world malaria map: Plasmodium falciparum endemicity in 2010. *Malar J* 2011; 10: 378 doi: 10.1186/1475-2875-10-378.
2. Malaria Atlas Project *P. falciparum* Cartographic code. https://github.com/malaria-atlas-project/mbgw-clean. Accessed December 2011.
3. Gilks WR, Spiegelhalter DJ. Markov Chain Monte Carlo in practice. Interdisciplinary Statistics. Boca Raton, Florida, U.S.A.: Chapman & Hall / CRC Press LLC, 1999.
4. Gelman A, Carlin JB, Stern HS. Bayesian data analysis. Texts in Statistical Science. Boca Raton, Florida, U.S.A.: Chapman & Hall / CRC Press LLC. 696 p, 1993.
5. Smith DL, Guerra CA, Snow RW, Hay SI. Standardizing estimates of the *Plasmodium falciparum* parasite rate. *Malar J* 2007; 6: 131.

6. Stein ML. Space-time covariance functions. *J Am Stat Assoc* 2005; 100: 310-321.

7. Azzalini A. A class of distributions which includes the normal ones. *Scand J Stat* 1985; 12: 171-178.

8. van Rossum G (2008) Python Programming Language - Official Website. Website: URL http://www.python.org.

9. Fonnesbeck C, Huard D, Patil AP. PyMC 2.0 User's Guide: installation and tutorial 2008. URL: http://www.trichech.us/pymc.

10. Patil A, Huard D, Fonnesbeck CJ. PyMC: Bayesian stochastic modelling in Python. *J Stat Softw* 2010; 35: e1000301.

11. Oliphant TE. Python for scientific computing. *Comput Sci Eng* 2007; 9: 10-20.

12. Haario H, Saksman E, Tamminen J. An adaptive Metropolis algorithm. *Bernoulli* 2001; 7: 223-242.

13. Hogg RV, Craig AT. *Introduction to Mathematical Statistics*. Upper Saddle River, New Jersey, U.S.A: Prentice Hall Inc. 2005; 564 p.

14. Guerra CA, Gikandi PW, Tatem AJ, Noor AM, Smith DL, *et al.* The limits and intensity of *Plasmodium falciparum* transmission: implications for malaria control and elimination worldwide. *PLoS Med* 2008; 5: e38.